GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 11984.5 Seconds
(without alignments)
11710.708 Million cell updates/sec

Title: US-09-610-313B-30
Perfect score: 2469
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

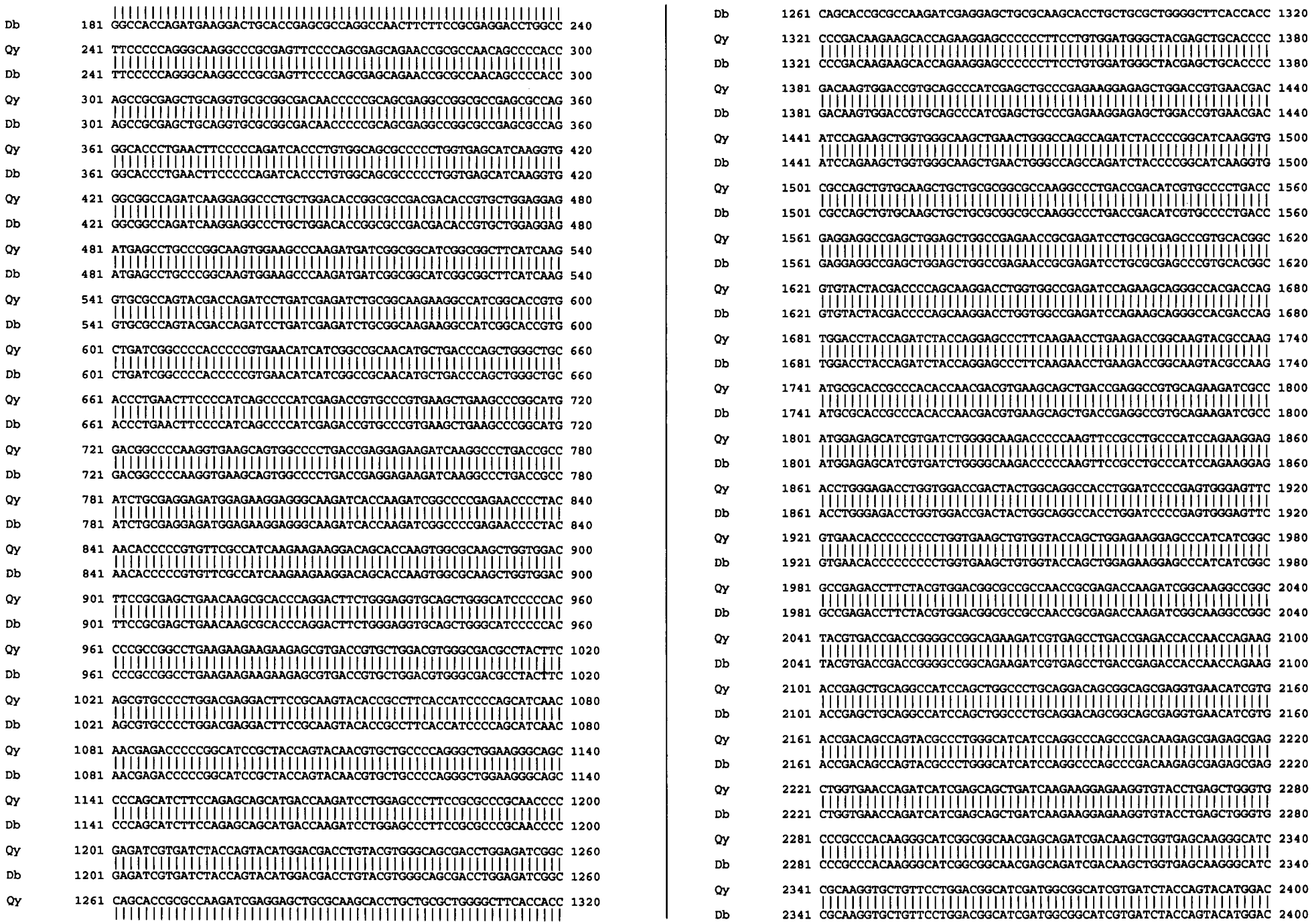
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2469	100.0	2469	6	AX455914	AX455914 Sequence
	2	2442.2	98.9	2463	6	AX455915	AX455915 Sequence
	3	2415.4	97.8	2457	6	AX455916	AX455916 Sequence
	4	2052	83.1	2312	6	BD263706	BD263706 Improved
	5	2052	83.1	2312	6	CQ870576	CQ870576 Sequence
	6	2052	83.1	2312	6	AR373389	AR373389 Sequence
	7	2051.2	83.1	9166	6	AX427930	AX427930 Sequence
	8	2042.8	82.7	9788	6	AX427936	AX427936 Sequence
	9	2028.8	82.2	9169	6	AX427931	AX427931 Sequence
	10	2027	82.1	9194	6	AX427925	AX427925 Sequence
	11	2027	82.1	12411	6	AX427927	AX427927 Sequence
	12	2025.4	82.0	9194	6	AX427926	AX427926 Sequence
	13	2025.2	82.0	2306	6	BD263704	BD263704 Improved
	14	2025.2	82.0	2306	6	CQ870574	CQ870574 Sequence
	15	2025.2	82.0	2306	6	AR373387	AR373387 Sequence
	16	2020.4	81.8	9785	6	AX427938	AX427938 Sequence
	17	2018.6	81.8	9189	6	AX427921	AX427921 Sequence
	18	2017	81.7	9167	6	AX427933	AX427933 Sequence
	19	2017	81.7	9170	6	AX427928	AX427928 Sequence
	20	2017	81.7	9782	6	AX427935	AX427935 Sequence
	21	2017	81.7	9783	6	AX427934	AX427934 Sequence
	22	2017	81.7	9790	6	CS078943	CS078943 Sequence
	23	2017	81.7	9792	6	AX427932	AX427932 Sequence
c	24	2017	81.7	36066	6	CS078956	CS078956 Sequence
	25	2015.6	81.6	4353	11	AF287353	AF287353 Synthetic
	26	1999	81.0	3015	6	AX455946	AX455946 Sequence
	27	1998.4	80.9	2300	6	BD263705	BD263705 Improved
	28	1998.4	80.9	2300	6	CQ870575	CQ870575 Sequence
	29	1998.4	80.9	2300	6	AR373388	AR373388 Sequence
	30	1982.4	80.3	3009	6	AX455987	AX455987 Sequence
	31	1969.4	79.8	4319	6	BD263639	BD263639 Improved
	32	1969.4	79.8	4319	6	CQ870498	CQ870498 Sequence
	33	1969.4	79.8	4319	6	AR373322	AR373322 Sequence
	34	1958.8	79.3	4352	11	AF287352	AF287352 Synthetic
	35	1958.8	79.3	4352	11	AY789509	AY789509 Synthetic
	36	1907	77.2	2577	6	AX457088	AX457088 Sequence
	37	1905.8	77.2	4343	6	AX306428	AX306428 Sequence
	38	1901.8	77.0	4341	6	AX306429	AX306429 Sequence
	39	1892.2	76.6	1977	6	AX455954	AX455954 Sequence
	40	1880.8	76.2	7344	6	CS078948	CS078948 Sequence
	41	1879.2	76.1	7897	6	AX427923	AX427923 Sequence
	42	1858	75.3	2305	6	BD263702	BD263702 Improved
	43	1858	75.3	2305	6	CQ870572	CQ870572 Sequence
	44	1858	75.3	2305	6	AR373385	AR373385 Sequence
	45	1837	74.4	1965	6	AX455952	AX455952 Sequence

ALIGNMENTS

RESULT 1
AX455914
LOCUS AX455914 2469 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 30 from Patent WO0204493.
ACCESSION AX455914
VERSION AX455914.1 GI:21714899
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof
JOURNAL Patent: WO 0204493-A 30 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES Location/Qualifiers
source 1. .2469
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975(+)"
ORIGIN

Query Match 100.0%; Score 2469; DB 6; Length 2469;
Best Local Similarity 100.0%; Pred. No. 6.7e-204;
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
```


Db 181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Qy 241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db 241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Qy 301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db 301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Qy 361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db 361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Qy 421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db 421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Qy 481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db 481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Qy 541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db 541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Qy 601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Qy 661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db 661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Qy 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Qy 781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db 781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Qy 841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db 841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy 961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db 961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy 1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Db 1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Qy 1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Db 1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Qy 1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Db 1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Qy 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Db 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Qy 1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Db 1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Qy 1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Db 1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Qy 1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Db 1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Qy 1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Db 1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Qy 1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Db 1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Qy 1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Db 1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Qy 1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Db 1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Qy 1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Db 1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Qy 1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Db 1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Qy 1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Db 1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Qy 2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Db 2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Qy 2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Db 2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Qy 2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Db 2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Qy 2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Db 2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Qy 2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Db 2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Qy 2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Db 2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

```
Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2461 GGTGAATTC 2469
             |||||||||
Db       2461 GGTGAATTC 2469


RESULT 2
AX455915
LOCUS       AX455915                2463 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 31 from Patent WO0204493.
ACCESSION   AX455915
VERSION     AX455915.1  GI:21714900
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 31 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1. .2463
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YM"
ORIGIN

  Query Match             98.9%;  Score 2442.2;  DB 6;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 1.4e-201;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              ||||||||||||||||  || |  ||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
```

```
Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy       1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGTGAATTC 2469
             |||||||||
Db       2455 GGTGAATTC 2463


RESULT 3
AX455916
LOCUS       AX455916               2457 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 32 from Patent WO0204493.
ACCESSION   AX455916
VERSION     AX455916.1  GI:21714901
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 32 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1. .2457
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YMWM"
ORIGIN

  Query Match             97.8%;  Score 2415.4;  DB 6;  Length 2457;
  Best Local Similarity   99.3%;  Pred. No. 2.7e-199;
  Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
```

```
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368
Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Db       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Db       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Db       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Db       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Db       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Db       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Db       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Db       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Db       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Db       1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Db       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Db       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Db       2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Db       2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Db       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Db       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Db       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388
Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Db       2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448
Qy       2461 GGTGAATTC 2469
Db       2449 GGTGAATTC 2457
```

```
RESULT 4
BD263706
LOCUS       BD263706               2312 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263706
VERSION     BD263706.1  GI:33073474
KEYWORDS    JP 2002533124-A/73.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 73 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/73
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH   Key
                     Location/Qualifiers
            FT   source          1. .2312
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
```

```
     source          1. .2312
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             83.1%;  Score 2052;  DB 6;  Length 2312;
  Best Local Similarity   93.6%;  Pred. No. 4.9e-168;
  Matches 2165;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||    || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              |||||||||||||||  |||||||||||||||||||  |||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              |||||||||| |||||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      ||||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
              |||||||||||| ||||| | | ||||||||| |||||||||| ||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
              |||| ||||| |||||||||||||||| ||||||||||||||    | ||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
              || | |||||||||||||||||| | ||||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
              || |||||||||| | |||||||||||||||||||||||||||||||||||||||| ||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
              |||||||||||||||| || |||||||||||||||| ||||||||||||||||||||||
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
              ||||||||||||||||    |  ||||||   |||||||||||||||||||||||| ||
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
              |||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db         961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
              ||||||||||    |  |||||||| ||||||||||||||||||||||||||||||||
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
              |||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
              |||||||| ||||||||||||||  |||||||||||||||||||||||| |||||||||
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
              |||||||  |  | ||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
              |||||   |||||||||||||  ||||||||||||||||||||||||||||||||||||
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
              |||||||||||||||  | ||||||||||||||||||||||||||||||||||||||
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
              ||||||||||||||||||    ||||||  ||||||||||||||||||||||||||||||
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
              |||||||||||||||||  |   ||  |||||||||||||||||||||||| ||||||||
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
              ||  |||||||||||||||||||||||||||||||||||||| ||||||||  |||||||
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
              |||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
              ||||||||  |||||||||||||||||||||||||||||||||||||  |  ||||||
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
              |  |||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
              |||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040
```

```
Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
             ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
             |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||| | |
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
             |||||||||||||||   |||||||||||||||||||||||| ||||||||||||||||||||
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
             |||||||||||||||||||||||||||||||
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312


RESULT 5
CQ870576
LOCUS       CQ870576                2312 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 84 from Patent EP1433851.
ACCESSION   CQ870576
VERSION     CQ870576.1  GI:52000092
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 84 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt(+)"
ORIGIN

  Query Match             83.1%;  Score 2052;  DB 6;  Length 2312;
  Best Local Similarity   93.6%;  Pred. No. 4.9e-168;
  Matches 2165;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              |||||||||||||||  |||||||||||||||||||  |||||||||| ||||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              |||||||||| |||||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      |||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
              |||||||||||| ||||| | | |||||||||| |||||||||| ||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||

Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
              |||| ||||| |||||||||||||||||| |||||||||||||||    | ||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
              || | |||||||||||||||||||| | ||||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
              || |||||||||| | ||||||||||||||||||||||||||||||||||||||| ||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
              |||||||||||||||| || |||||||||||||||| ||||||||||||||||||||||
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
              |||||||||||||||| | |||||| |||||||||||||||||||||||||||||| ||
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
              |||||||||||  |||||||| ||||||||||||||||||||||||||||||||||||||
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
              |||||||||||||||||||||||||||||||| |||||||||||||||||||| ||||||
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
              |||||||| ||||||||||||||   |||||||||||||||||||||||| |||||||||
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
              ||||||| |  | |||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440
```

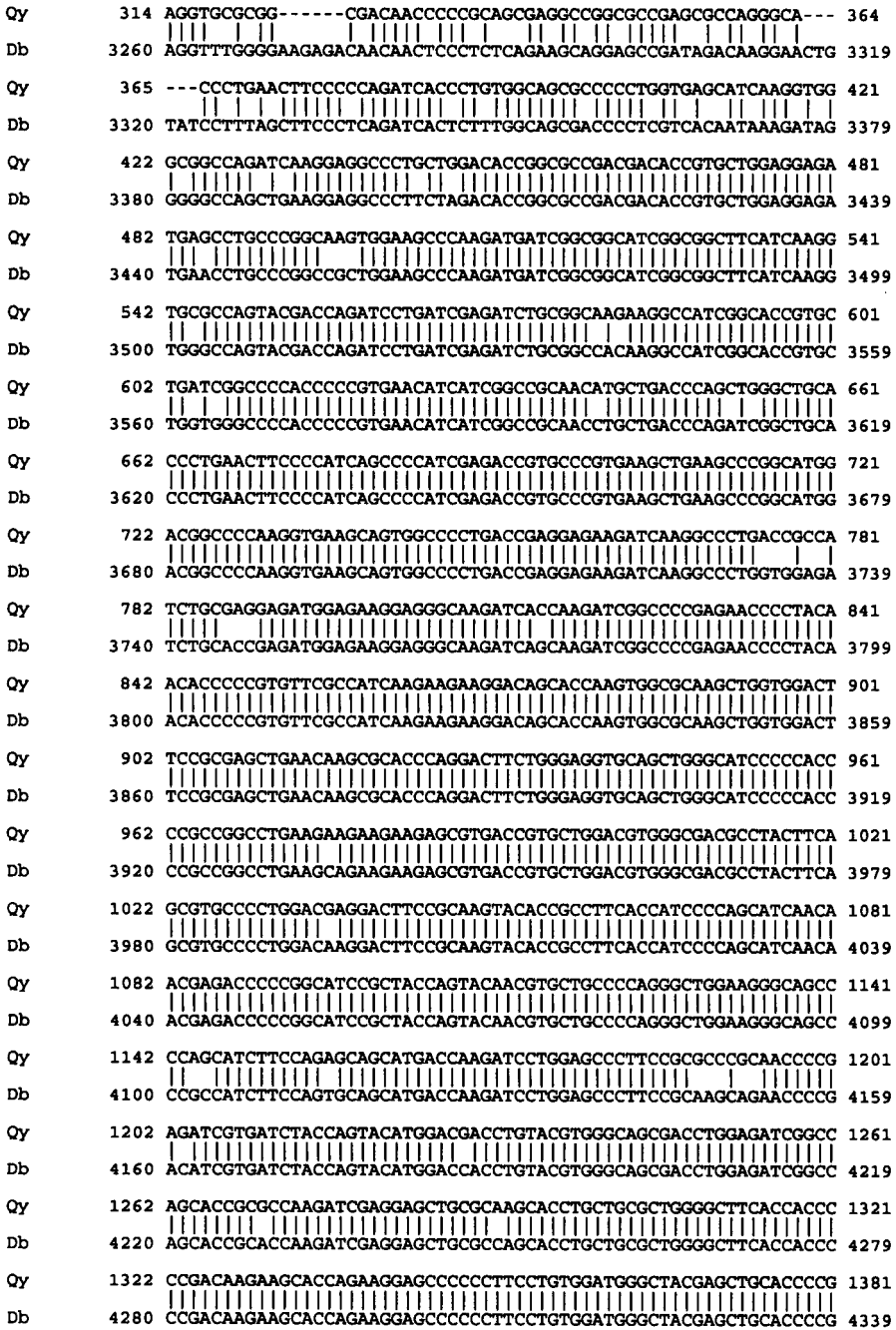
```
Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

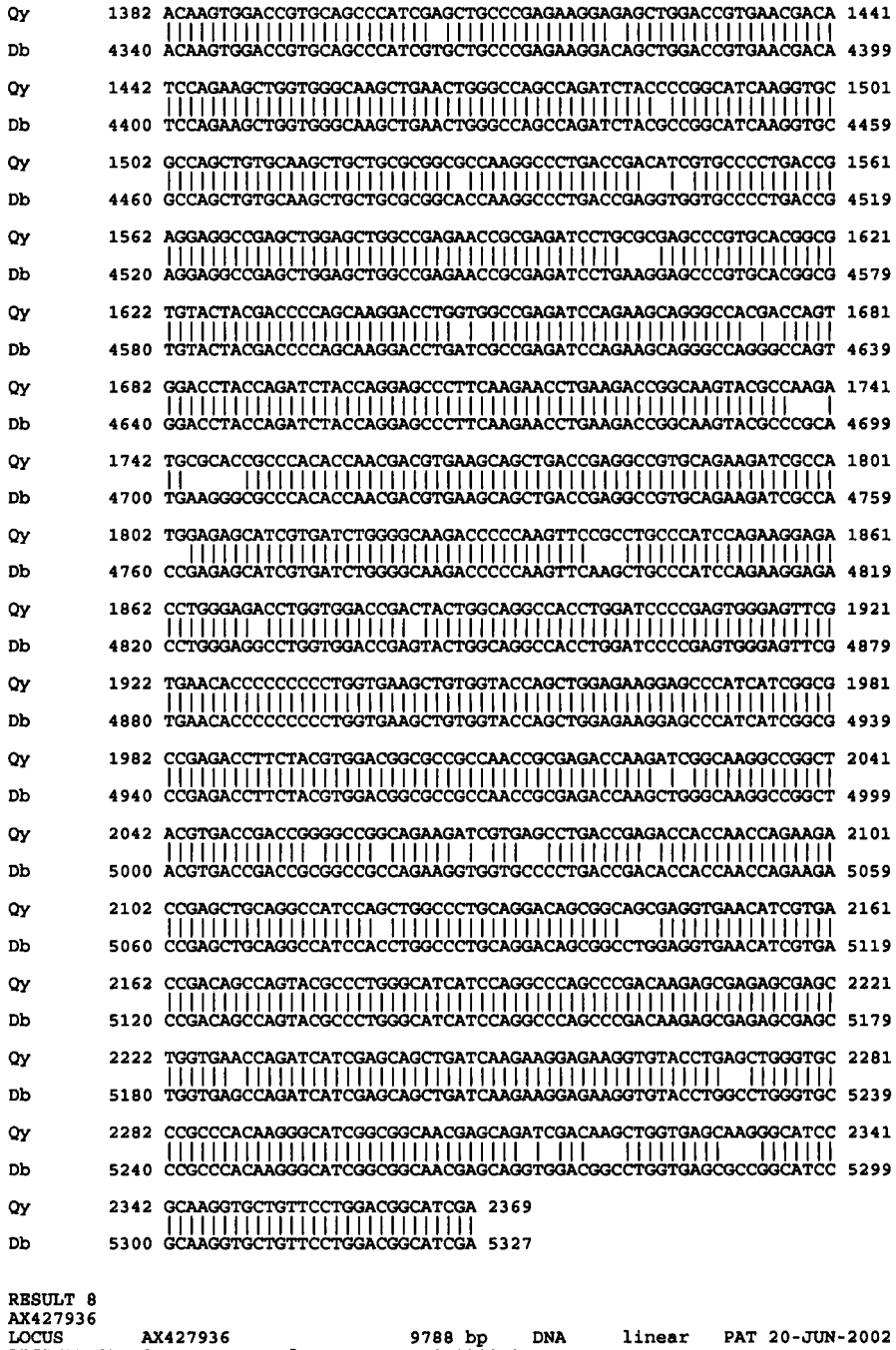
```
RESULT 6
AR373389
LOCUS       AR373389               2312 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 84 from patent US 6602705.
ACCESSION   AR373389
VERSION     AR373389.1  GI:40075492
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 84 05-AUG-2003;
            Chiron Corporation; Emeryville, CA
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

Query Match 83.1%; Score 2052; DB 6; Length 2312;
Best Local Similarity 93.6%; Pred. No. 4.9e-168;
Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
```

```
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297

Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357

Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417

Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477

Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537

Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597

Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657

Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717

Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777

Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837

Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1897

Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957

Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017

Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077

Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137

Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197

Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257

Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317

Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377

Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437

Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469

Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312


RESULT 7
AX427930
LOCUS       AX427930               9166 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 168 from Patent WO0232943.
ACCESSION   AX427930
VERSION     AX427930.1  GI:21538017
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 168 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9166
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             83.1%;  Score 2051.2;  DB 6;  Length 9166;
  Best Local Similarity   92.6%;  Pred. No. 3.6e-168;
  Matches 2192;  Conservative    0;  Mismatches  163;  Indels   13;  Gaps    3;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193

Db       3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253

Db       3141 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3199

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313

Db       3200 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3259
```





RESULT 8
AX427936
LOCUS AX427936 9788 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 174 from Patent WO0232943.

```
ACCESSION   AX427936
VERSION     AX427936.1  GI:21538023
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 174 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9788
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match           82.7%;  Score 2042.8;  DB 6;  Length 9788;
  Best Local Similarity   91.8%;  Pred. No. 1.9e-167;
  Matches 2197;  Conservative    0;  Mismatches  177;  Indels   18;  Gaps    3;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       3081 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       3141 AGGACTGCACCGAGCGACAGGCTAA------TAGGGAAGATCTGGCCTTCCCACAAGGGA 3194

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       3195 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3254

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db       3255 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3314

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db       3315 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3374

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db       3375 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3434

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db       3435 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3494

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db       3495 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3554

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db       3555 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3614

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db       3615 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3674

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db       3675 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3734

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db       3735 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3794

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db       3795 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3854

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db       3855 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3914

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db       3915 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3974

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db       3975 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4034

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db       4035 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4094

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db       4095 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4154

Qy       1202 AGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1261
Db       4155 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4214

Qy       1262 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1321
Db       4215 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4274

Qy       1322 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1381
Db       4275 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4334

Qy       1382 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1441
Db       4335 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4394

Qy       1442 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1501
Db       4395 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4454

Qy       1502 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1561
Db       4455 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4514

Qy       1562 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1621
Db       4515 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4574

Qy       1622 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1681
Db       4575 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4634

Qy       1682 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1741
Db       4635 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4694

Qy       1742 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1801
Db       4695 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4754

Qy       1802 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1861
```

```
Db       4755 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4814

Qy       1862 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1921
             |||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       4815 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4874

Qy       1922 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1981
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4875 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4934

Qy       1982 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2041
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       4935 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4994

Qy       2042 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2101
             |||||||||||| ||||| |||||| | |||  ||||||||| ||||||||||||||||||
Db       4995 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5054

Qy       2102 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2161
             |||||||||||||||||| |||||||||||||||||||   ||||||||||||||||||||
Db       5055 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5114

Qy       2162 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2221
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5115 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5174

Qy       2222 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2281
             |||||| |||||||||||||||||||||||||||||||||||||||  ||||||||
Db       5175 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5234

Qy       2282 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2341
             |||||||||||||||||||||||||||||| | |||  |||||||||  |||||||
Db       5235 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5294

Qy       2342 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
             |||||||||||||||||||||||||||||    | |   | |    || || |
Db       5295 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5346


RESULT 9
AX427931
LOCUS       AX427931               9169 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 169 from Patent WO0232943.
ACCESSION   AX427931
VERSION     AX427931.1  GI:21538018
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 169 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9169
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             82.2%;  Score 2028.8;  DB 6;  Length 9169;
  Best Local Similarity   91.7%;  Pred. No. 3.1e-166;
  Matches 2194;  Conservative    0;  Mismatches  182;  Indels   16;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||||||||||||        | |||||||| |||||||||| |||||||||
Db       2967 TGGCCGAGGCCATGAGCCAGGTGCAGCACACCAACATCATGATGCAGCGCGGCAACTTCA 3026

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||  ||||||||   |||||||||||||||||||||||||||| | |||||||

Db       3027 AGGGCCAGAAGCGCAT---CAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3083

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db       3084 ACTGCCGCGCCCTGCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 3143

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              |||||||||||||||| ||||| || || ||   | || || |||||||||| || || |
Db       3144 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3202

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||| | || ||  |     ||||||| | | |||||||||||||||  | ||||| |
Db       3203 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3262

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
              |||| | ||      | ||||| ||| |   || || || |||||  | || || |
Db       3263 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3322

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
                 || | ||||||  |||||||| || |||||||| |||||  || |  || ||| | |
Db       3323 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3382

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
              | |||||  | |||||||||| || ||||||||||||||||||||||||||||||||
Db       3383 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3442

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
              ||| ||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db       3443 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3502

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
              || |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       3503 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3562

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
              || | |||||||||||||||||||||||||||||||||||| |||||||||| | |||||||
Db       3563 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3622

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3623 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3682

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||    | |
Db       3683 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3742

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
              |||||   |||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       3743 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3802

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3803 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3862

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3863 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3922

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       3923 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3982

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       3983 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4042

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4043 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4102

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
              ||  |||||||||| |||||||||||||||||||||||||||    |  |||||||
Db       4103 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4162
```

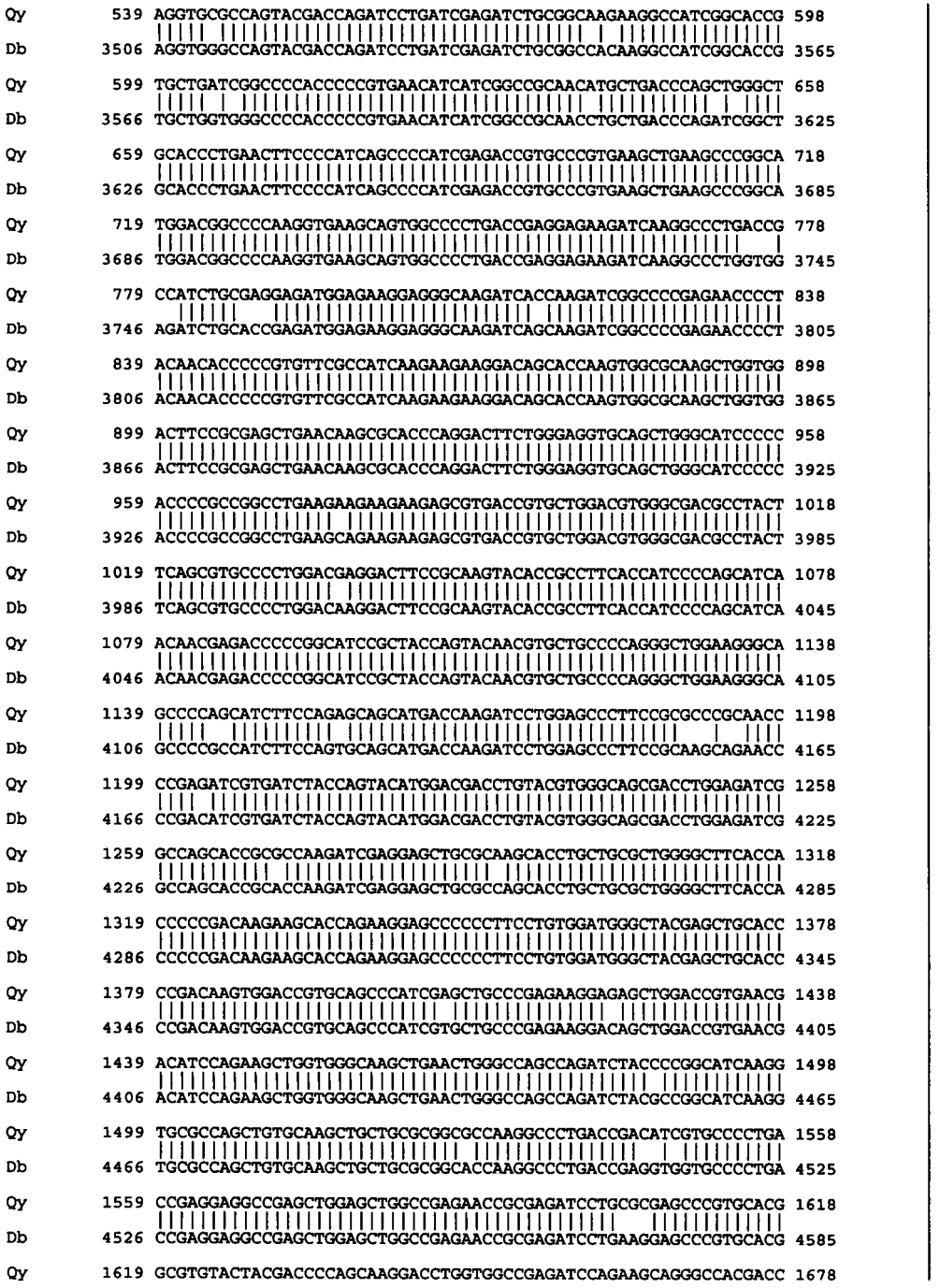
```
Qy       1202 AGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1261
Db       4163 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4222

Qy       1262 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1321
Db       4223 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4282

Qy       1322 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1381
Db       4283 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4342

Qy       1382 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1441
Db       4343 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4402

Qy       1442 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1501
Db       4403 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4462

Qy       1502 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1561
Db       4463 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4522

Qy       1562 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1621
Db       4523 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4582

Qy       1622 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1681
Db       4583 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4642

Qy       1682 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1741
Db       4643 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4702

Qy       1742 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1801
Db       4703 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4762

Qy       1802 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1861
Db       4763 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4822

Qy       1862 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1921
Db       4823 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4882

Qy       1922 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1981
Db       4883 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4942

Qy       1982 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2041
Db       4943 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 5002

Qy       2042 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2101
Db       5003 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5062

Qy       2102 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2161
Db       5063 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5122

Qy       2162 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2221
Db       5123 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5182

Qy       2222 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2281
Db       5183 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5242

Qy       2282 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2341
Db       5243 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5302

Qy       2342 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db       5303 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5354
```

RESULT 10
AX427925

```
LOCUS       AX427925                9194 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 163 from Patent WO0232943.
ACCESSION   AX427925
VERSION     AX427925.1  GI:21538012
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 163 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9194
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN
```

Query Match 82.1%; Score 2027; DB 6; Length 9194;
Best Local Similarity 91.6%; Pred. No. 4.4e-166;
Matches 2194; Conservative 0; Mismatches 185; Indels 16; Gaps 4;

```
Qy         14 TGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAACT 70
Db       2967 TGGCCGAGGCCATGAGCCAGGTGACCAACAGCGCCACCATCATGATGCAGCGCGGCAACT 3026

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGGCCACATCGCCC 130
Db       3027 TCCGCAACCAGCGCAAGATCGTGAAGTGCTTCAACTGCGGCAAGGAGGGGCCACACCGCCC 3086

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
Db       3087 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 3146

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
Db       3147 TGAAGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAG 3205

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
Db       3206 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 3265

Qy        311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
Db       3266 TTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 3325

Qy        365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
Db       3326 CTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGA 3385

Qy        419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
Db       3386 TAGGGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGG 3445

Qy        479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
Db       3446 AGATGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 3505
```




```
RESULT 11
AX427927
LOCUS       AX427927               12411 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 165 from Patent WO0232943.
ACCESSION   AX427927
VERSION     AX427927.1  GI:21538014
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 165 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .12411
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
```

/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match 82.1%; Score 2027; DB 6; Length 12411;
Best Local Similarity 91.6%; Pred. No. 4e-166;
Matches 2194; Conservative 0; Mismatches 185; Indels 16; Gaps 4;

```
Qy      14 TGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAACT 70
           ||||||||||||||||||||   ||| |||||||| |||| |||||||||||| ||||||
Db    2967 TGGCCGAGGCCATGAGCCAGGTGACCAACAGCGCCACCATCATGATGCAGCGCGGCAACT 3026

Qy      71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
           ||      ||         ||| | ||||||||||||||||||||||||||||||||||
Db    3027 TCCGCAACCAGCGCAAGATCGTGAAGTGCTTCAACTGCGGCAAGGAGGGCCACACCGCCC 3086

Qy     131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3087 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 3146

Qy     191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
           |||||||||||||||||| ||||| || || ||  | || || |||||||||||| || |
Db    3147 TGAAGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAG 3205

Qy     251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
           | |||||| | || ||  |     ||||||| | | |||||||||||||||||  | ||||
Db    3206 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 3265

Qy     311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
           | |||||  | ||      | |||||  ||| |    || || || |||||  | || || |
Db    3266 TTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 3325

Qy     365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
                 || | | |||||| |||||||| || |||||||| ||||| || |   || |||
Db    3326 CTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGA 3385

Qy     419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
           | || |||||| | |||||||||| || ||||||||||||||||||||||||||||||||
Db    3386 TAGGGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGG 3445

Qy     479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
           |||||| ||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db    3446 AGATGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 3505

Qy     539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
           ||||| |||||||||||||||||||||||||||||||||||||| | ||||||||||||||
Db    3506 AGGTGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCG 3565

Qy     599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCT 658
           ||||| | |||||||||||||||||||||||||||||||| |||||||||| | ||||
Db    3566 TGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCT 3625

Qy     659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3626 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 3685

Qy     719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||   |
Db    3686 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGG 3745

Qy     779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
             ||||||   |||||||||||||||||||||||| |||||||||||||||||||||||||
Db    3746 AGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCT 3805

Qy     839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3806 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 3865

Qy     899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3866 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 3925

Qy     959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    3926 ACCCCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 3985

Qy    1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    3986 TCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 4045

Qy    1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4046 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 4105

Qy    1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
           |||||  |||||||||| ||||||||||||||||||||||||||||||  |  ||||
Db    4106 GCCCCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACC 4165

Qy    1199 CCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 1258
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4166 CCGACATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 4225

Qy    1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    4226 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 4285

Qy    1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4286 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 4345

Qy    1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438
           |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db    4346 CCGACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 4405

Qy    1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4406 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 4465

Qy    1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558
           |||||||||||||||||||||||||||||| |||||||||||||| | ||||||||||
Db    4466 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGA 4525

Qy    1559 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1618
           |||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db    4526 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 4585

Qy    1619 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1678
           |||||||||||||||||||||||||||||| | |||||||||||||||||||| | ||
Db    4586 GCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCC 4645

Qy    1679 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4646 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 4705

Qy    1739 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1798
             |||      ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4706 GCATGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 4765

Qy    1799 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1858
           |||  |||||||||||||||||||||||||||||||||||  |||||||||||||||||
Db    4766 CCACCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGG 4825

Qy    1859 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1918
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    4826 AGACCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 4885

Qy    1919 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1978
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4886 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 4945

Qy    1979 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038
           |||||||||||||||||||||||||||||||||||||||||||||||||| | ||||||||
Db    4946 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCG 5005

Qy    2039 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2098
           |||||||||||||||| |||| |||||| | ||| ||||||||| ||||||||||||||
Db    5006 GCTACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGA 5065
```

```
Qy       2099 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2158
             |||||||||||||||||||||| |||||||||||||||||||||   |||||||||||||
Db       5066 AGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG 5125

Qy       2159 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2218
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5126 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 5185

Qy       2219 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2278
             |||||||||| ||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       5186 AGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGG 5245

Qy       2279 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2338
             |||||||||||||||||||||||||||||||||||| | |||   |||||||||   ||||
Db       5246 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCA 5305

Qy       2339 TCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
             |||||||||||||||||||||||||||||||    | |   | |    || || |
Db       5306 TCCGCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5360


RESULT 12
AX427926
LOCUS       AX427926               9194 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 164 from Patent WO0232943.
ACCESSION   AX427926
VERSION     AX427926.1  GI:21538013
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 164 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9194
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             82.0%;  Score 2025.4;  DB 6;  Length 9194;
  Best Local Similarity   91.6%;  Pred. No. 6e-166;
  Matches 2193;  Conservative    0;  Mismatches  186;  Indels   16;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAACATCCTGATGCAGCGCAGCAACT 70
              ||||||||||||||||||||   ||| |||||||| |||| |||||||||| ||||||
Db       2967 TGGCCGAGGCCATGAGCCAGGTGACCAACAGCGCCACCATCATGATGCAGCGCGGCAACT 3026

Qy         71 TCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGGCCACATCGCCC 130
              ||      ||         ||| | ||||||||||||||||||||||||||||||| |||||
Db       3027 TCCGCAACCAGCGCAAGATCGTGAAGTGCTTCAACTGCGGCAAGGAGGGGCCACACCGCCC 3086

Qy        131 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3087 GCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGA 3146

Qy        191 TGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGG 250
              |||||||||||||||||| ||||| || || ||  | || || |||||||||||| || |
Db       3147 TGAAGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAG 3205

Qy        251 GCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGC 310
              | |||||| | || ||  |     ||||||| | | ||||||||||||||||  | ||||
Db       3206 GGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGC 3265

Qy        311 TGCAGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA 364
              | |||||  | ||      | ||||| ||| |   || || || |||||  | || || |
Db       3266 TTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAA 3325

Qy         365 ------CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGG 418
                   || | | |||||| |||||||| || |||||||| ||||| || |  || |||
Db       3326 CTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGA 3385

Qy         419 TGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGG 478
               | || |||||| | |||||||||||| || |||||||||||||||||||||||||||||
Db       3386 TAGGGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGG 3445

Qy         479 AGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538
               |||||| ||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db       3446 AGATGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 3505

Qy         539 AGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598
               ||||| |||||||||||||||||||||||||||||||||| | ||||||||||||||||
Db       3506 AGGTGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCG 3565

Qy         599 TGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACCATGCTGACCCAGCTGGGCT 658
               ||||| | |||||||||||||||||||||||||||||||||||||| |||||||||| | ||||
Db       3566 TGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCT 3625

Qy         659 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 718
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3626 GCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCA 3685

Qy         719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCG 778
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||    |
Db       3686 TGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGG 3745

Qy         779 CCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCT 838
                 ||||||   |||||||||||||||||||||||| |||||||||||||||||||||||
Db       3746 AGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCT 3805

Qy         839 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 898
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3806 ACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGG 3865

Qy         899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 958
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3866 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCC 3925

Qy         959 ACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 1018
               |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       3926 ACCCCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACT 3985

Qy        1019 TCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 1078
               |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       3986 TCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCA 4045

Qy        1079 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 1138
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4046 ACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCA 4105

Qy        1139 GCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACC 1198
               |||||  |||||||||| ||||||||||||||||||||||||||||||    | ||||
Db       4106 GCCCCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACC 4165

Qy        1199 CCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCG 1258
               |||| |||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       4166 CCGACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCG 4225

Qy        1259 GCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA 1318
               |||||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||
Db       4226 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA 4285

Qy        1319 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 1378
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4286 CCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACC 4345

Qy        1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACG 1438
               |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db       4346 CCGACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACG 4405
```

```
Qy       1439 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1498
Db       4406 ACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGG 4465

Qy       1499 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGA 1558
Db       4466 TGCGCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGA 4525

Qy       1559 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACG 1618
Db       4526 CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACG 4585

Qy       1619 GCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACC 1678
Db       4586 GCGTGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCC 4645

Qy       1679 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
Db       4646 AGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCC 4705

Qy       1739 AGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 1798
Db       4706 GCATGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG 4765

Qy       1799 CCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGG 1858
Db       4766 CCACCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGG 4825

Qy       1859 AGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 1918
Db       4826 AGACCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGT 4885

Qy       1919 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 1978
Db       4886 TCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCG 4945

Qy       1979 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038
Db       4946 GCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCG 5005

Qy       2039 GCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGA 2098
Db       5006 GCTACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGA 5065

Qy       2099 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCG 2158
Db       5066 AGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCG 5125

Qy       2159 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 2218
Db       5126 TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCG 5185

Qy       2219 AGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGG 2278
Db       5186 AGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGG 5245

Qy       2279 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCA 2338
Db       5246 TGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCA 5305

Qy       2339 TCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db       5306 TCCGCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5360
```

```
RESULT 13
BD263704
LOCUS       BD263704               2306 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263704
VERSION     BD263704.1  GI:33073472
KEYWORDS    JP 2002533124-A/71.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 71 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/71
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH   Key
                      Location/Qualifiers
            FT   source          1. .2306
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN
```

Query Match 82.0%; Score 2025.2; DB 6; Length 2306;
Best Local Similarity 93.3%; Pred. No. 1e-165;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
```

```
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600
Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660
Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720
Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780
Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840
Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900
Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960
Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020
Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074
Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db       1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134
Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db       1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1194
Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db       1195 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1254
Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db       1255 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1314
Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db       1315 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1374
Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db       1375 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1434
Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db       1435 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1494
Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db       1495 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1554
Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db       1555 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1614
Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db       1615 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1674
Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db       1675 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1734
Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db       1735 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1794
Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db       1795 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1854
Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db       1855 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1914
Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db       1915 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1974
Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db       1975 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2034
Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db       2035 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2094
Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
Db       2095 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2154
Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
Db       2155 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2214
Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
Db       2215 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2274
Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
Db       2275 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2306
```

```
RESULT 14
CQ870574
LOCUS       CQ870574               2306 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 82 from Patent EP1433851.
ACCESSION   CQ870574
VERSION     CQ870574.1  GI:52000090
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 82 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt.YM"
ORIGIN
```

Query Match 82.0%; Score 2025.2; DB 6; Length 2306;
Best Local Similarity 93.3%; Pred. No. 1e-165;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

```
Qy   170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db     1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG  60

Qy   230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db    61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy   290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db   121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy   344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db   181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy   398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db   241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy   458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db   301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy   518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db   361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy   578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db   421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy   638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db   481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy   698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db   541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy   758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db   601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy   818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db   661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy   878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db   721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy   938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db   781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy   998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db   841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy  1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db   901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy  1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db   961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy  1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237

Db  1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074

Qy  1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
Db  1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134

Qy  1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
Db  1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1194

Qy  1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db  1195 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1254

Qy  1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
Db  1255 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1314

Qy  1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
Db  1315 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1374

Qy  1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
Db  1375 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1434

Qy  1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
Db  1435 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1494

Qy  1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
Db  1495 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1554

Qy  1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
Db  1555 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1614

Qy  1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837
Db  1615 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1674

Qy  1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897
Db  1675 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1734

Qy  1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957
Db  1735 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1794

Qy  1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017
Db  1795 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1854

Qy  2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077
Db  1855 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1914

Qy  2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137
Db  1915 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1974

Qy  2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197
Db  1975 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2034

Qy  2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db  2035 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2094

Qy  2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317
```

```
Db       2095 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2154

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377
             |||||||||||||   |||||||||||||||||||||||| ||||||||||||||||||||
Db       2155 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2214

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2274

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469
             |||||||||||||||||||||||||||||||
Db       2275 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2306


RESULT 15
AR373387
LOCUS       AR373387                2306 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 82 from patent US 6602705.
ACCESSION   AR373387
VERSION     AR373387.1  GI:40075490
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 82 05-AUG-2003;
            Chiron Corporation; Emeryville, CA
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             82.0%;  Score 2025.2;  DB 6;  Length 2306;
  Best Local Similarity   93.3%;  Pred. No. 1e-165;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              |||||||||||||||   |||||||||||||||||||||||  |||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              |||||||||| |||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||| ||      |||||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
              |||||||||||| ||||| | | |||||||||| |||||||||| ||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
              |||| ||||| |||||||||||||||| ||||||||||||||   | ||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
              || | |||||||||||||||||| | ||||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
              || |||||||||| | |||||||||||||||||||||||||||||||||||||| ||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
              |||||||||||||||| || |||||||||||||||| ||||||||||||||||||||||
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
              |||||||||||||||||   | ||||||   |||||||||||||||||||||||||| ||
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
              |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACG 1237
              ||||||||||   | |||||||| ||||||||||||||      || | ||||||||
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCA------GGCCCCCCTGTACG 1074

Qy       1238 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1297
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||| ||||||
Db       1075 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1134

Qy       1298 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1135 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1194

Qy       1358 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
              |||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db       1195 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1254

Qy       1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1477
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1314

Qy       1478 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1537
              |||||||| ||||||||||||||   |||||||||||||||||||||||| |||||||||
Db       1315 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1374

Qy       1538 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1597
              |||||||  |  | ||||||||||||||||||||||||||||||||||||||||||||
Db       1375 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1434

Qy       1598 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1657
              |||||  |||||||||||||  ||||||||||||||||||||||||||||||||||||
Db       1435 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1494

Qy       1658 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1717
              |||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db       1495 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1554

Qy       1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1777
```

```
Db       1555 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1614

Qy       1778 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1837

Db       1615 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1674

Qy       1838 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897

Db       1675 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1734

Qy       1898 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1957

Db       1735 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1794

Qy       1958 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2017

Db       1795 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1854

Qy       2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2077

Db       1855 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1914

Qy       2078 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2137

Db       1915 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1974

Qy       2138 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2197

Db       1975 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2034

Qy       2198 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2257

Db       2035 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2094

Qy       2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2317

Db       2095 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2154

Qy       2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2377

Db       2155 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2214

Qy       2378 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2437

Db       2215 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2274

Qy       2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2469

Db       2275 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2306
```

Search completed: December 30, 2005, 19:16:16
Job time : 11995.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 1299.83 Seconds
(without alignments)
12659.489 Million cell updates/sec

Title: US-09-610-313B-30
Perfect score: 2469
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2469	100.0	2469	6	ABL39959	Abl39959 Synthetic
2	2469	100.0	2469	12	ADM73764	Adm73764 HIV-1 pol
3	2457	99.5	2457	8	ACA03548	Aca03548 Synthetic
4	2457	99.5	2457	10	ADC13266	Adc13266 DNA of HI
5	2442.2	98.9	2463	6	ABL39960	Abl39960 Synthetic
6	2442.2	98.9	2463	12	ADM73765	Adm73765 HIV-1 pol
7	2434.6	98.6	2457	8	ACA03547	Aca03547 Synthetic
8	2434.6	98.6	2457	10	ADC13265	Adc13265 DNA of HI
9	2415.4	97.8	2457	6	ABL39961	Abl39961 Synthetic
10	2415.4	97.8	2457	12	ADM73766	Adm73766 HIV-1 pol
11	2401.8	97.3	2445	8	ACA03546	Aca03546 Synthetic
12	2401.8	97.3	2445	10	ADC13264	Adc13264 DNA of HI
13	2394.8	97.0	3930	10	ADC13230	Adc13230 DNA of HI
14	2393.2	96.9	3930	10	ADC13231	Adc13231 DNA of HI
15	2393.2	96.9	3930	10	ADC13232	Adc13232 DNA of HI
16	2393.2	96.9	5184	8	ACA03591	Aca03591 Synthetic
17	2393.2	96.9	5184	10	ADC13279	Adc13279 DNA of HI
18	2362.8	95.7	3531	10	ADC13234	Adc13234 DNA of HI
19	2361.2	95.6	3537	10	ADC13236	Adc13236 DNA of HI
20	2360.2	95.6	5145	8	ACA03521	Aca03521 Synthetic
21	2360.2	95.6	5145	10	ADC13233	Adc13233 DNA of HI
22	2350.2	95.2	3538	10	ADC13235	Adc13235 DNA of HI
23	2349.4	95.2	3624	8	ACA03550	Aca03550 Synthetic
24	2349.4	95.2	3624	10	ADC13268	Adc13268 DNA of HI
25	2301.8	93.2	3607	8	ACA03551	Aca03551 Synthetic
26	2301.8	93.2	3607	10	ADC13269	Adc13269 DNA of HI
27	2283.6	92.5	3597	8	ACA03549	Aca03549 Synthetic
28	2283.6	92.5	3597	10	ADC13267	Adc13267 DNA of HI
29	2149.6	87.1	2472	8	ACA03543	Aca03543 Synthetic
30	2149.6	87.1	2472	8	ACC78507	Acc78507 HIV p2Pol
31	2121.2	85.9	2466	8	ACA03542	Aca03542 Synthetic
32	2121.2	85.9	2466	8	ACC78506	Acc78506 HIV p2Pol
33	2094.4	84.8	2460	8	ACA03541	Aca03541 Synthetic
34	2094.4	84.8	2460	8	ACC78505	Acc78505 HIV p2Pol
35	2093.4	84.8	3564	8	ACC78488	Acc78488 HIV GagPo
36	2093.4	84.8	3564	8	ACC78489	Acc78489 HIV GagPo
37	2092.8	84.8	4716	8	ACA03522	Aca03522 Synthetic
38	2092.8	84.8	4716	10	ADC13238	Adc13238 DNA of HI
39	2089	84.6	3999	8	ACC78484	Acc78484 HIV GagCo
40	2087.4	84.5	3999	8	ACC78485	Acc78485 HIV GagCo
41	2087.4	84.5	3999	8	ACC78486	Acc78486 HIV GagCo
42	2087.4	84.5	5283	8	ACA03584	Aca03584 Synthetic
43	2087.4	84.5	5283	8	ACC78529	Acc78529 HIV TatRe
44	2087.2	84.5	4713	8	ACA03592	Aca03592 Synthetic
45	2087.2	84.5	4713	10	ADC13280	Adc13280 DNA of HI

ALIGNMENTS

RESULT 1
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 8; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef

```
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 2469;  DB 6;  Length 2469;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-298;
  Matches 2469;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
```

```
Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2461 GGTGAATTC 2469
             |||||||||
Db       2461 GGTGAATTC 2469
```

RESULT 2
ADM73764
ID ADM73764 standard; DNA; 2469 BP.
XX
AC ADM73764;
XX
DT 03-JUN-2004 (first entry)
XX
DE HIV-1 polynucleotide #7.
XX
KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW HIV type C protein; immunostimulant.
XX
OS Human immunodeficiency virus 1.
XX
PN US2003223961-A1.
XX
PD 04-DEC-2003.
XX
PF 05-JUL-2001; 2001US-00899575.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (MEGE/) MEGEDE J Z.
PA (BARN/) BARNETT S W.
PA (ENGE/) ENGELBRECHT S.
PA (RENS/) RENSBURG E J V.
XX
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;
XX
DR WPI; 2004-060515/06.
XX
PT New expression cassette comprising a polynucleotide sequence encoding an
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT immunization, generating of packaging cell lines or in producing HIV Type
PT C proteins.
XX
PS Claim 1; SEQ ID NO 30; 160pp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV Type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
XX
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 100.0%; Score 2469; DB 12; Length 2469;
Best Local Similarity 100.0%; Pred. No. 8.3e-298;
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Qy       2461 GGTGAATTC 2469
Db       2461 GGTGAATTC 2469
```

RESULT 3
ACA03548
ID ACA03548 standard; DNA; 2457 BP.
XX
AC ACA03548;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #31.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.

XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 36; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 99.5%; Score 2457; DB 8; Length 2457;
Best Local Similarity 100.0%; Pred. No. 2.6e-296;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1927 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

```
RESULT 4
ADC13266
ID   ADC13266 standard; DNA; 2457 BP.
XX
AC   ADC13266;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 42; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;
```

Query Match 99.5%; Score 2457; DB 10; Length 2457;
Best Local Similarity 100.0%; Pred. No. 2.6e-296;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
```

```
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600
Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660
Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720
Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780
Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840
Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900
Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960
Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020
Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080
Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 5
ABL39960
ID ABL39960 standard; DNA; 2463 BP.
XX
AC ABL39960;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975YM SEQ ID NO:31.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.

XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 9; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 98.9%; Score 2442.2; DB 6; Length 2463;
Best Local Similarity 99.6%; Pred. No. 1.7e-294;
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              ||||||||||||||||||      || |  ||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
```

```
Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGTGAATTC 2469
             |||||||||
Db       2455 GGTGAATTC 2463
```

```
RESULT 6
ADM73765
ID   ADM73765 standard; DNA; 2463 BP.
XX
AC   ADM73765;
XX
DT   03-JUN-2004  (first entry)
XX
DE   HIV-1 polynucleotide #8.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
PD   04-DEC-2003.
XX
PF   05-JUL-2001; 2001US-00899575.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (MEGE/) MEGEDE J Z.
PA   (BARN/) BARNETT S W.
PA   (ENGE/) ENGELBRECHT S.
PA   (RENS/) RENSBURG E J V.
XX
PI   Megede JZ,  Barnett SW,  Engelbrecht S,  Rensburg EJV;
XX
DR   WPI; 2004-060515/06.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding an
PT   HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT   immunization, generating of packaging cell lines or in producing HIV Type
PT   C proteins.
XX
PS   Claim 1; SEQ ID NO 31; 160pp; English.
XX
CC   The invention relates to an expression cassette comprising a
CC   polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC   also relates to a recombinant expression system for use in a host cell
CC   comprising an expression cassette, where the polynucleotide sequence
CC   further comprises control elements capable of driving expression in the
CC   selected host cell, a cell comprising an expression cassette where the
CC   polynucleotide sequence further comprises control elements compatible
CC   with the expression in the cell and a composition for generating an
CC   immunological response, comprising an expression cassette. The expression
CC   cassette and the methods of the invention are useful in eliciting an
CC   immune response, in DNA immunisation, in generation of packaging cell
CC   lines and in producing HIV Type C proteins. This sequence represents an
CC   HIV-1 polynucleotide of the invention.
XX
SQ   Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;
```

```
  Query Match             98.9%;  Score 2442.2;  DB 12;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 1.7e-294;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              ||||||||||||||||      || |  ||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy       1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGTGAATTC 2469
              |||||||||
Db       2455 GGTGAATTC 2463
```

RESULT 7
ACA03547
ID ACA03547 standard; DNA; 2457 BP.

XX
AC ACA03547;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #30.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 35; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 98.6%; Score 2434.6; DB 8; Length 2457;
Best Local Similarity 99.6%; Pred. No. 1.5e-293;
Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

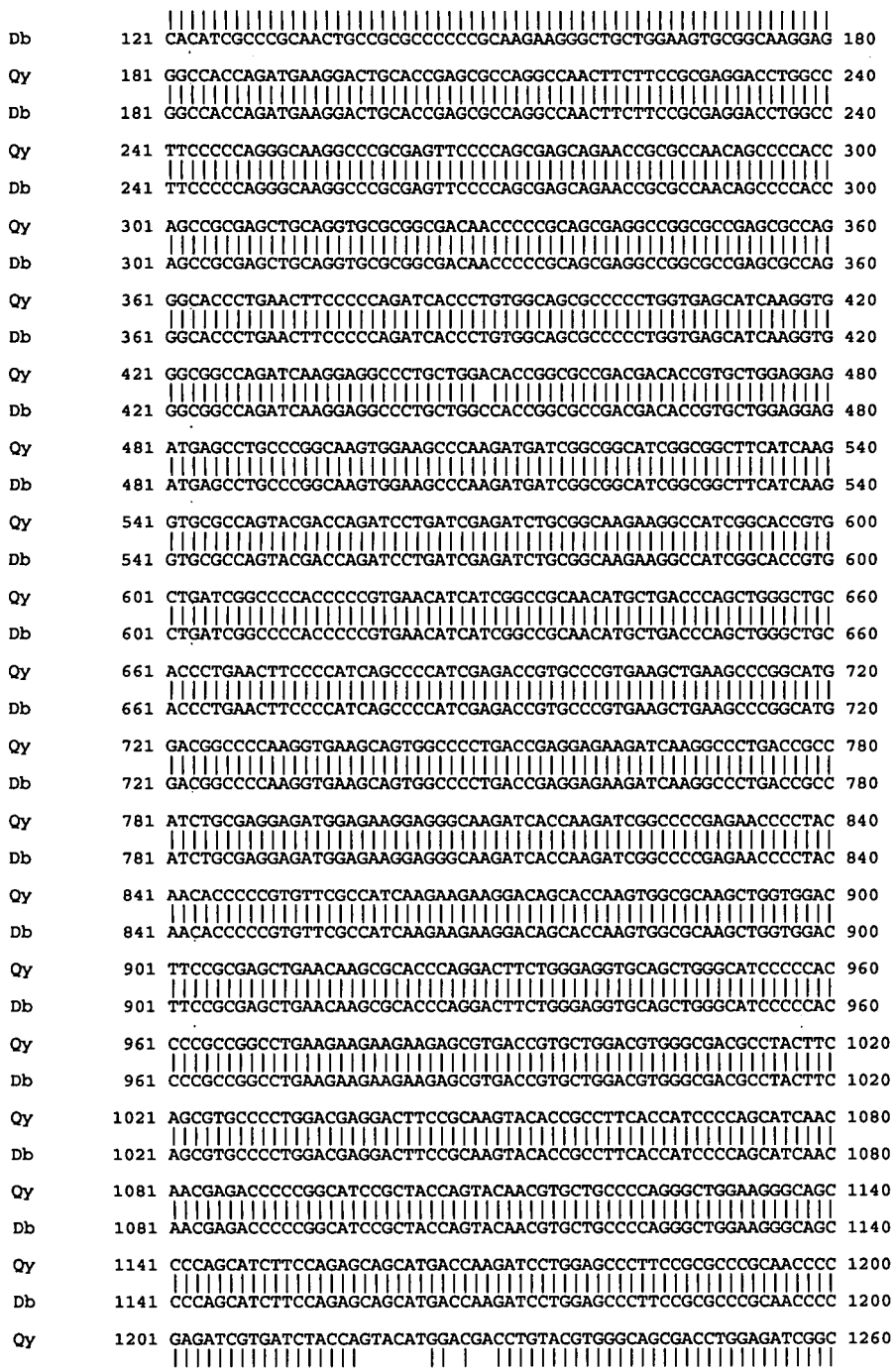
```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              |||||||||||||||||| || | ||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
```

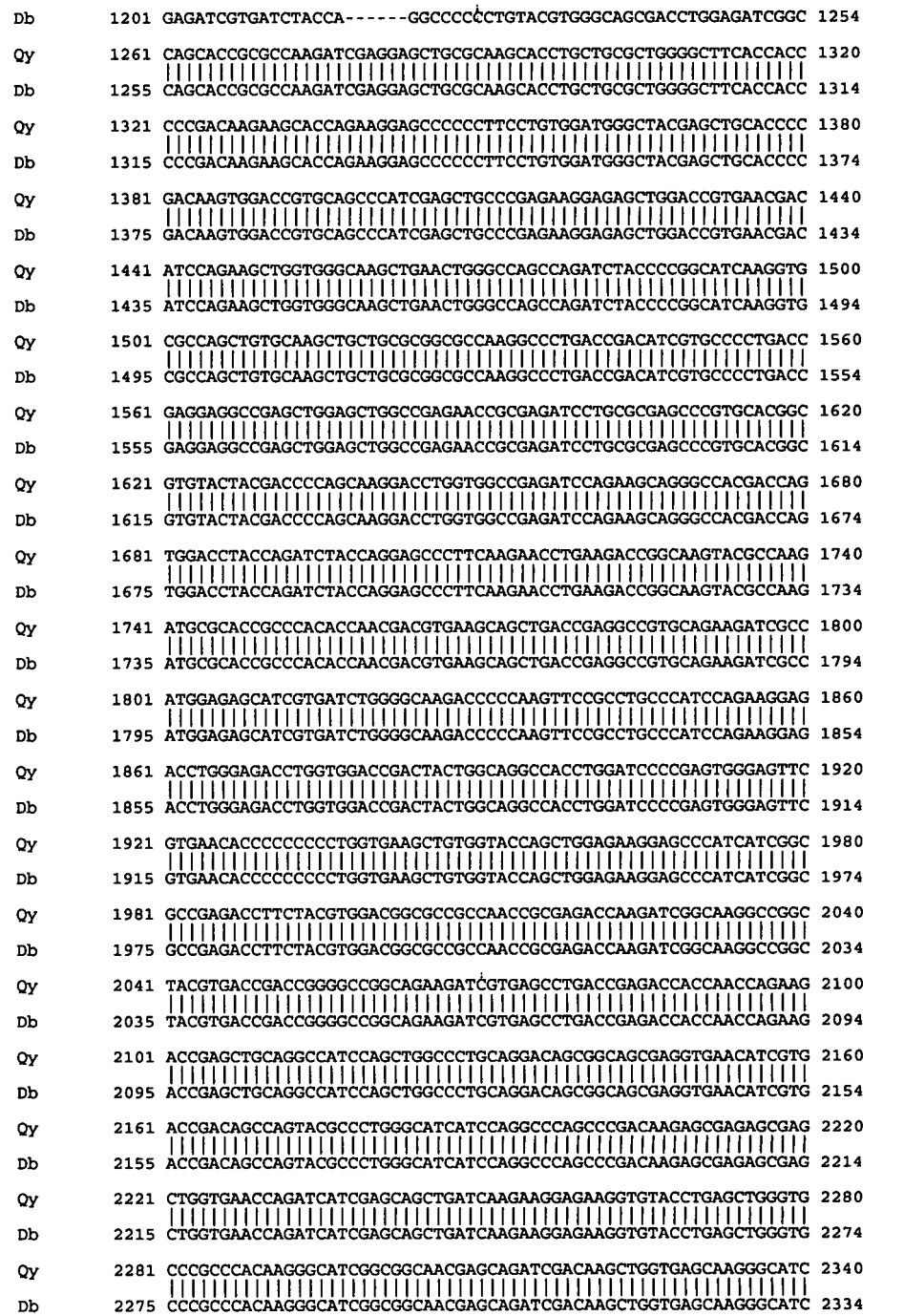
```
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914

Qy       1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGT 2463
              |||
Db       2455 GGT 2457
```

RESULT 8
ADC13265
ID ADC13265 standard; DNA; 2457 BP.
XX
AC ADC13265;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2Pol-opt-YM_C SEQ ID NO 44.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 41; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match   98.6%; Score 2434.6; DB 10; Length 2457;
  Best Local Similarity 99.6%; Pred. No. 1.5e-293;
  Matches 2453; Conservative   0; Mismatches   4; Indels   6; Gaps   1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
```





```
Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454

Qy       2461 GGT 2463
             |||
Db       2455 GGT 2457


RESULT 9
ABL39961
ID   ABL39961 standard; DNA; 2457 BP.
XX
AC   ABL39961;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975YMWM SEQ ID NO:32.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
PT   New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT   applications including DNA immunization or generation of packaging cell
PT   lines, particularly in gene therapy.
XX
PS   Claim 1; Fig 10; 233pp; English.
XX
CC   The present invention describes expression cassettes comprising a
CC   polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC   type C polypeptides. The expression cassettes comprise any of the HIV
CC   type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

  Query Match            97.8%;  Score 2415.4;  DB 6;  Length 2457;
  Best Local Similarity  99.3%;  Pred. No. 3.7e-291;
  Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
```

```
Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
             |||||||||||||||||      || |  ||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
             |||||||||||||||||||||||||||||||||||    ||       ||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908

Qy       1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088

Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448

Qy       2461 GGTGAATTC 2469
             |||||||||
Db       2449 GGTGAATTC 2457
```

RESULT 10
ADM73766
ID ADM73766 standard; DNA; 2457 BP.
XX
AC ADM73766;
XX
DT 03-JUN-2004 (first entry)
XX
DE HIV-1 polynucleotide #9.
XX
KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW HIV type C protein; immunostimulant.
XX
OS Human immunodeficiency virus 1.
XX
PN US2003223961-A1.
XX
PD 04-DEC-2003.
XX
PF 05-JUL-2001; 2001US-00899575.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (MEGE/) MEGEDE J Z.
PA (BARN/) BARNETT S W.
PA (ENGE/) ENGELBRECHT S.
PA (RENS/) RENSBURG E J V.
XX
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;
XX
DR WPI; 2004-060515/06.
XX
PT New expression cassette comprising a polynucleotide sequence encoding an
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT immunization, generating of packaging cell lines or in producing HIV Type
PT C proteins.
XX
PS Claim 1; SEQ ID NO 32; 160pp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV Type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
XX

SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 97.8%; Score 2415.4; DB 12; Length 2457;
Best Local Similarity 99.3%; Pred. No. 3.7e-291;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
              ||||||||||||||||  ||  |||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254

Qy       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314

Qy       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
              ||||||||||||||||||||||||||||||||||||  ||      ||||||||||||||
Db       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368

Qy       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428

Qy       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488

Qy       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548

Qy       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608

Qy       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668

Qy       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728

Qy       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788

Qy       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848

Qy       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908

Qy       1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968

Qy       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028

Qy       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
```

```
Qy       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148

Qy       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208

Qy       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268

Qy       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328

Qy       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388

Qy       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448

Qy       2461 GGTGAATTC 2469
             |||||||||
Db       2449 GGTGAATTC 2457
```

RESULT 11
ACA03546
ID ACA03546 standard; DNA; 2445 BP.
XX
AC ACA03546;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #29.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 34; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

```
  Query Match             97.3%;  Score 2401.8;  DB 8;  Length 2445;
  Best Local Similarity   99.2%;  Pred. No. 1.8e-289;
  Matches 2438;  Conservative    0;  Mismatches    7;  Indels   12;  Gaps    2;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
```

```
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966

Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026

Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086

Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146

Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206

Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266

Db       1201 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1254

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326

Db       1255 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1314

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386

Db       1315 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1368

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446

Db       1369 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1428

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506

Db       1429 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1488

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566

Db       1489 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1548

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626

Db       1549 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1608

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686

Db       1609 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1668

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746

Db       1669 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1728

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806

Db       1729 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1788

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866

Db       1789 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1848

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926

Db       1849 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1908

Qy       1927 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986

Db       1909 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1968

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046

Db       1969 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2028

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106

Db       2029 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2088

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166

Db       2089 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2148

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226

Db       2149 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2208

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286

Db       2209 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2268

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346

Db       2269 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2328

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406

Db       2329 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2388

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463

Db       2389 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 12
ADC13264
ID ADC13264 standard; DNA; 2445 BP.
XX
AC ADC13264;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2Pol-opt-YMWM_C SEQ ID NO 43.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 40; 301pp; English.

```
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match             97.3%;  Score 2401.8;  DB 10;  Length 2445;
  Best Local Similarity   99.2%;  Pred. No. 1.8e-289;
  Matches 2438;  Conservative    0;  Mismatches    7;  Indels   12;  Gaps    2;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1266
              ||||||||||| || | ||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1254

Qy       1267 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1314

Qy       1327 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1386
              |||||||||||||||||||||||||||||||| || |||||||||||||||||||||||
Db       1315 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1368

Qy       1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1428

Qy       1447 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1488

Qy       1507 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1548

Qy       1567 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1608

Qy       1627 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1668

Qy       1687 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1728

Qy       1747 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1788

Qy       1807 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1848

Qy       1867 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1908
```

```
Qy       1927 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1968

Qy       1987 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2028

Qy       2047 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2029 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2088

Qy       2107 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2089 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2148

Qy       2167 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2149 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2208

Qy       2227 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2209 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2268

Qy       2287 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2269 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2328

Qy       2347 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2329 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2388

Qy       2407 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2389 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 13
ADC13230
ID ADC13230 standard; DNA; 3930 BP.
XX
AC ADC13230;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagComplPolmut_C SEQ ID NO 9.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 6; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 97.0%; Score 2394.8; DB 10; Length 3930;
Best Local Similarity 99.2%; Pred. No. 1.2e-288;
Matches 2431; Conservative 0; Mismatches 7; Indels 12; Gaps 2;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
```

```
Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db      2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db      2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy      1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
Db      2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy      1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
Db      2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy      1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
Db      2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854

Qy      1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
Db      2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy      1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
Db      2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy      1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
Db      2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy      1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
Db      3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094

Qy      1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
Db      3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy      1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
Db      3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy      1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
Db      3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy      1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
Db      3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy      1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
Db      3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy      1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
Db      3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy      1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
Db      3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy      2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
Db      3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574

Qy      2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
Db      3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy      2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
Db      3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy      2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
Db      3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy      2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
Db      3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy      2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
Db      3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy      2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
Db      3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 14
ADC13231
ID   ADC13231 standard; DNA; 3930 BP.
XX
AC   ADC13231;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutAtt_C SEQ ID NO 10.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
```

PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 7; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;

Query Match 96.9%; Score 2393.2; DB 10; Length 3930;
Best Local Similarity 99.2%; Pred. No. 1.9e-288;
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;

```
Qy        14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy        74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy       134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1607 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy       194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy       254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy       314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy       374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy       434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy       494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy       554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy       614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy       674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy       734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy      1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
             ||||   || |  ||||||||||||||||||||||||||||||||||||||||||||||
Db      2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740

Qy      1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800

Qy      1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
             ||||||||||||||||||||||   ||       ||||||||||||||||||||||||||
Db      2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854

Qy      1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914

Qy      1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974

Qy      1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034

Qy      1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094

Qy      1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154

Qy      1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214

Qy      1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274

Qy      1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
```

```
Db       3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334

Qy       1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933

Db       3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394

Qy       1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993

Db       3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454

Qy       1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053

Db       3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514

Qy       2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113

Db       3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574

Qy       2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173

Db       3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634

Qy       2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233

Db       3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694

Qy       2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293

Db       3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754

Qy       2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353

Db       3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814

Qy       2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413

Db       3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874

Qy       2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463

Db       3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924


RESULT 15
ADC13232
ID   ADC13232 standard; DNA; 3930 BP.
XX
AC   ADC13232;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutIna_C SEQ ID NO 11.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 8; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;

  Query Match             96.9%;  Score 2393.2;  DB 10;  Length 3930;
  Best Local Similarity   99.2%;  Pred. No. 1.9e-288;
  Matches 2430;  Conservative    0;  Mismatches    8;  Indels   12;  Gaps    2;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193

Db       1607 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253

Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313

Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373

Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433

Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493

Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553

Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613

Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673

Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
```

```
Db      2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206
Qy       734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db      2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db      2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db      2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626
Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
Qy      1214 ACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 1273
Db      2687 ACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCA 2740
Qy      1274 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 1333
Db      2741 AGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGC 2800
Qy      1334 ACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCG 1393
Db      2801 ACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCG 2854
Qy      1394 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
Db      2855 TGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914
Qy      1454 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513
Db      2915 TGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974
Qy      1514 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 1573
Db      2975 AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGC 3034
Qy      1574 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 1633
Db      3035 TGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACC 3094
Qy      1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 1693
Db      3095 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGA 3154
Qy      1694 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753
Db      3155 TCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 3214
Qy      1754 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 1813
Db      3215 ACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCG 3274
Qy      1814 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 1873
Db      3275 TGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCT 3334
Qy      1874 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 1933
Db      3335 GGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCC 3394
Qy      1934 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
Db      3395 CCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454
Qy      1994 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053
Db      3455 ACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3514
Qy      2054 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 2113
Db      3515 GGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGG 3574
Qy      2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 2173
Db      3575 CCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGT 3634
Qy      2174 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 2233
Db      3635 ACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGA 3694
Qy      2234 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 2293
Db      3695 TCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGG 3754
Qy      2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
Db      3755 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814
Qy      2354 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 2413
Db      3815 TCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGG 3874
Qy      2414 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2463
Db      3875 GCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

Search completed: December 30, 2005, 08:56:53
Job time : 1308.83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:41 ; Search time 8718.85 Seconds
(without alignments)
13249.138 Million cell updates/sec

Title: US-09-610-313B-30
Perfect score: 2469
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

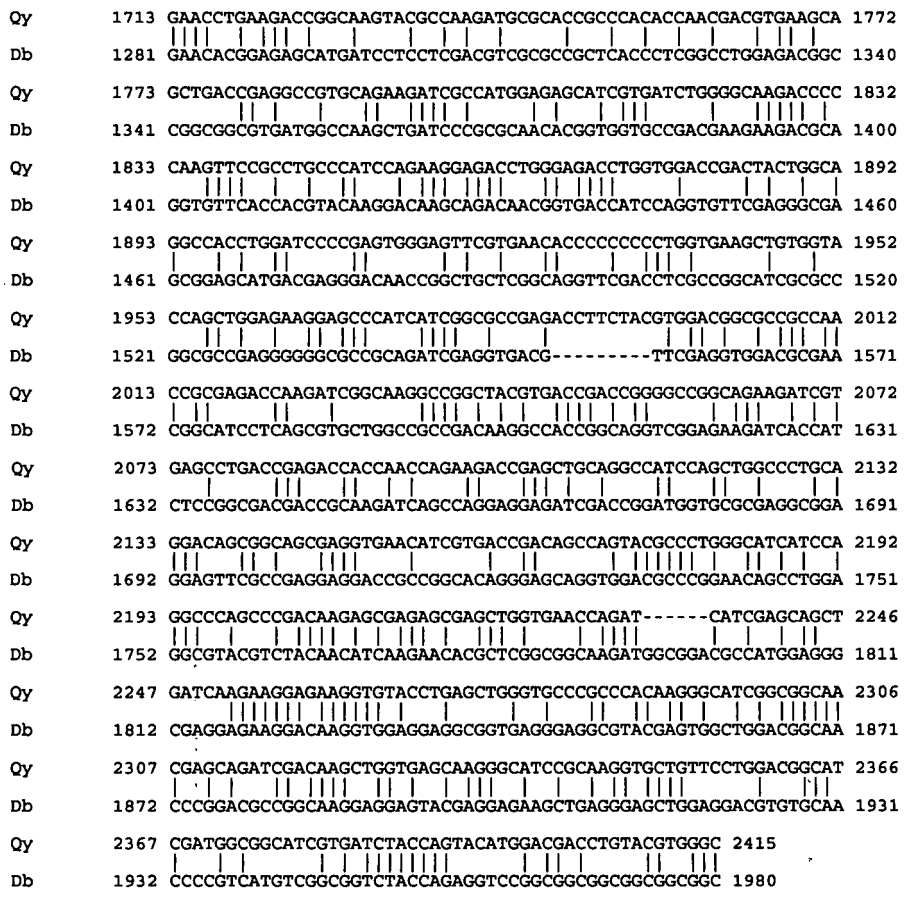
SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	136.4	5.5	330	10	CL293849	CL293849 02S0349-0
	2	97.8	4.0	2031	10	CL974989	CL974989 OsIFCC042
	3	97.2	3.9	2886	10	CL967755	CL967755 OsIFCC015
	4	91.4	3.7	1509	10	CL959255	CL959255 OsIFCC002
	5	90.8	3.7	2598	4	AY103647	AY103647 Zea mays
	6	89.4	3.6	743	10	CZ247380	CZ247380 AIAA-aaf3
	7	86.6	3.5	1398	10	CL961989	CL961989 OsIFCC006
	8	85.6	3.5	951	3	BM321451	BM321451 rockefell
c	9	85	3.4	869	7	CK159167	CK159167 FGAS04056
	10	84	3.4	3069	10	CL973991	CL973991 OsIFCC025
	11	83.8	3.4	1941	10	CL971508	CL971508 OsIFCC021
	12	82.6	3.3	1781	8	CX099623	CX099623 RECm0260
c	13	80.8	3.3	1060	10	CW922203	CW922203 EDCAR29TR
	14	80.6	3.3	892	10	CZ216254	CZ216254 AIAA-aaf2
	15	80.2	3.2	1132	3	BM320864	BM320864 rockefell
	16	79.6	3.2	1165	3	BM320900	BM320900 rockefell
	17	79	3.2	867	3	BM321430	BM321430 rockefell
	18	78.4	3.2	11691	10	CL962901	CL962901 OsIFCC008
	19	78.2	3.2	1962	10	CL961326	CL961326 OsIFCC005
	20	78.2	3.2	2853	10	CL974397	CL974397 OsIFCC025
	21	78	3.2	1485	10	CL970981	CL970981 OsIFCC020
	22	76.6	3.1	1550	3	BM321022	BM321022 rockefell
	23	76.6	3.1	2559	10	CL982027	CL982027 OsIFSC046
	24	76.2	3.1	1401	10	CL962721	CL962721 OsIFCC038
	25	75.6	3.1	671	6	CA093222	CA093222 SCCCCL300
	26	75.4	3.1	2299	4	AY106831	AY106831 Zea mays
	27	75.2	3.0	854	7	CK777127	CK777127 964131 MA
	28	75	3.0	862	3	BM321023	BM321023 rockefell
	29	75	3.0	3822	10	CL972913	CL972913 OsIFCC023
	30	74.6	3.0	914	9	BZ568300	BZ568300 pacs2-164
	31	74.6	3.0	2010	10	CL962831	CL962831 OsIFCC008
	32	74.6	3.0	2313	10	CL982362	CL982362 OsIFSC047
	33	74.6	3.0	2682	10	CL969033	CL969033 OsIFCC017
	34	74.4	3.0	3134	4	AY109500	AY109500 Zea mays
	35	74.2	3.0	1680	10	CL982770	CL982770 OsIFSC049
	36	74	3.0	1290	10	CL972679	CL972679 OsIFCC023
	37	74	3.0	1386	11	DQ045165	DQ045165 Homo sapi
	38	74	3.0	2072	4	CR603312	CR603312 full-leng
c	39	73.8	3.0	889	7	CK159613	CK159613 FGAS04107
c	40	73.6	3.0	757	9	CC678788	CC678788 OGWII71TV
	41	73.6	3.0	2028	10	CL979437	CL979437 OsIFCC033
	42	73.2	3.0	853	3	BM321393	BM321393 rockefell
	43	73	3.0	1689	10	CL972373	CL972373 OsIFCC022
c	44	72.8	2.9	566	3	BM587428	BM587428 170006873
	45	72.8	2.9	1575	10	CL979927	CL979927 OsIFCC045

ALIGNMENTS

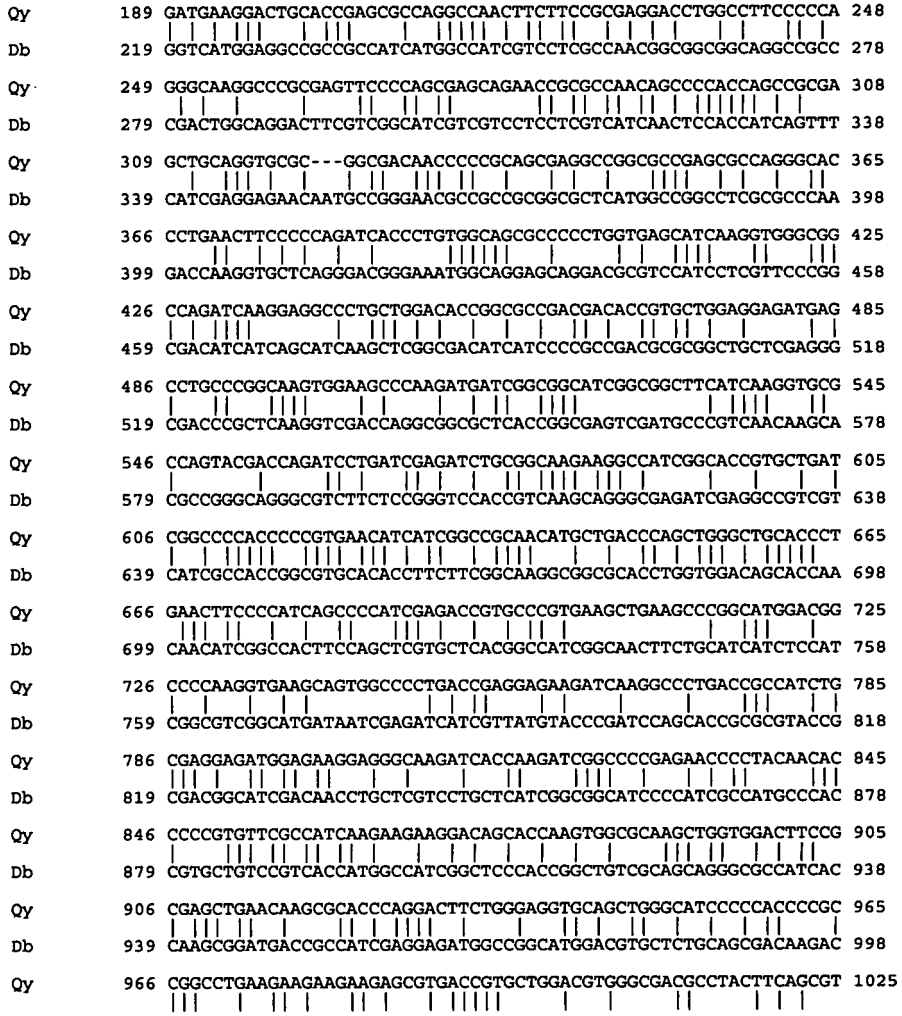
RESULT 1
CL293849/c
LOCUS CL293849 330 bp DNA linear GSS 12-FEB-2004
DEFINITION 02S0349-08A1-C03 UniformMu MuTAIL Library Zea mays genomic clone 02S0349-08A1-C03, genomic survey sequence.
ACCESSION CL293849
VERSION CL293849.1 GI:42541978
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 330)
AUTHORS Latshaw,S., Tan,B-.C., Settles,A.M. and McCarty,D.R.
TITLE Sequence tagged transposon insertions from the UniformMu maize population
JOURNAL Unpublished (2003)
COMMENT Contact: Donald R. McCarty
Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drm@ufl.edu
Sequence flanking probable Mu insertion site in UniformMu line: 02S0349-08, Primer set: A
Class: transposon insertion site.
FEATURES Location/Qualifiers
source 1. .330
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="02S0349-08A1-C03"
/clone_lib="UniformMu MuTAIL Library"
/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo assymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sepharose 400 spin columns and cloned into the TOPO PCR4 vector."
ORIGIN

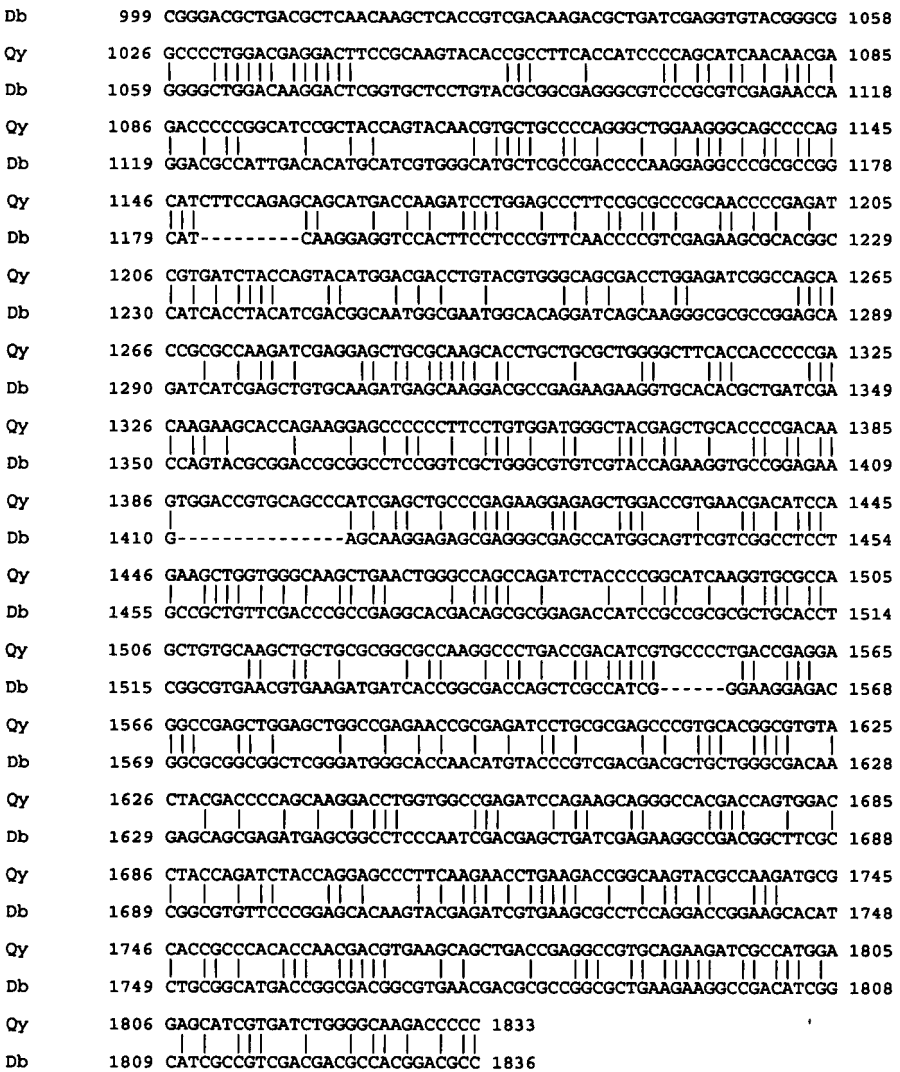
```
  Query Match             5.5%;  Score 136.4;  DB 10;  Length 330;
  Best Local Similarity   66.4%;  Pred. No. 2.8e-14;
  Matches  211;  Conservative    0;  Mismatches  106;  Indels    1;  Gaps    1;

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              || || ||| |  | || || ||  | || || || || || || || || ||  | || ||
Db        329 GGGGGGCAGCTGGAAGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAA 270

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||   || | || |  ||||| || || ||||| || || || || || || |||||
Db        269 ATGACTTTGACAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAA 210

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||  | ||||| || ||| | |  || || ||||| ||  | || || || || || |||
Db        209 GTAAGACAGTATGATCAGGTACCCATAGAAATCTGTGGGCATAAAGCTATAGGTACGGTA 150

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              |  | || || || || || ||||| || ||  | ||  || |||| ||| | || |||
Db        149 TTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGG-TGC 91

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||| | || || ||||| || || || || || || || || ||  | ||||| || |||
Db         90 ACCTTAAATTTTCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATG 31

Qy        721 GACGGCCCCAAGGTGAAG 738
              || ||||| || || | |
Db         30 GATGGCCCAAAAGTAAGG 13


RESULT 2
CL974989
LOCUS       CL974989               2031 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC042895 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL974989
VERSION     CL974989.1  GI:52404497
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2031)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .2031
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             4.0%;  Score 97.8;  DB 10;  Length 2031;
  Best Local Similarity   42.6%;  Pred. No. 2.7e-07;
  Matches  787;  Conservative    0;  Mismatches 1032;  Indels   30;  Gaps    4;

Qy        576 CGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCG 635
              ||||| || |  |    ||  ||   |  | |   |  ||  ||  | |||||||  |

Db        153 CGGCACGACGTACTCGTGCGTCGGAGTCTACCGGAACGGCCATGTCGAGATCATCGCCAA 212

Qy        636 CAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGT 695
              | ||  |  || || | |     | |   |  |  |  || |  |  | ||||    |
Db        213 CGACCAGGGGAACCGGATCACGCCCTCGTGGGTCGCCTTCACCGACGGCGGCGAGCGGCT 272

Qy        696 GCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGA 755
                 ||   ||            |  ||  |  ||||   | ||| |      |     ||
Db        273 CATCGGTGAGGCCGCCAAGAACCAGGCGGCGGCCAACCCGGAGCGGACCATCTACGACGC 332

Qy        756 GGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCAC 815
              ||  | |||  | ||  |     |  |  ||  ||| || ||   ||    ||| |
Db        333 CAAGCGGCTCATCGGCCGGCAGTTCTCCGACGCCGAGGTGCAGCGCGACATGAAGCTGCT 392

Qy        816 CAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAG 875
              | ||| |  ||  |||  ||  ||  ||     |  | || ||  ||||||| |
Db        393 GCCGTTCGCCGTCGTCGACCGGAACGGCAAGCCGCACGTGCGCGTCGAGGTGAAGGACGG 452

Qy        876 CACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTG 935
              |  |  |  |           | ||| ||  |  |  |  | || ||  ||   ||
Db        453 CGACGTGCG------------CGTGTTCAGCCCGGAGGAGGTGAGCGCCATGGTGCTCAC 500

Qy        936 GGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGT 995
              |  | ||  ||  |     |     |  |  ||||  ||||   |      |||   |
Db        501 GCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCACCCGCGCCGTCGTCAC 560

Qy        996 GCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACAC 1055
              |    |   |  |  | ||      ||       | |||||  | ||     | |
Db        561 CGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGGACGCCGGCGTCATCGC 620

Qy       1056 CGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGT 1115
              ||   ||||| ||  | |||||| |||||| || ||| |  |  |  |  |||  | |
Db        621 CGGGCTCACCGTCGACCGCATCATCAACGAGCCCACCGCCGCCGCCATCGCCTACGGCAT 680

Qy       1116 GCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175
                 |||||   ||  | |  |  |||||  |  | ||    || | ||
Db        681 CGACAAGAAGGGCGCCGAGAAGAACGTCCTCGTCTTCGACCTCGGCGGCGGCACGTTCGA 740

Qy       1176 GGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTA 1235
              |  | ||| ||||  | ||  ||   |  |      ||    ||   ||| |
Db        741 CGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTGCCACCAACGGCGACAC 800

Qy       1236 CGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCA 1295
              |     | ||| |  ||| ||| ||| | || |  |  |  |  |  |  | || |
Db        801 CCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACTTCGTCAAGGTCATCCG 860

Qy       1296 CCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCT 1355
              || |  || | |  ||  || |||| ||||  |     |  |  ||| || |  |
Db        861 CCGGAAGCACGGCCGCGACATCACCGGCGACGCGCGCGCGCTGGGCAAGCTCCGCCGCGA 920

Qy       1356 GTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCT---GCC 1412
              |||   | ||  | |||      | |  |||   |  |  |  || | |||||  |   | |
Db        921 GTGCGAGCGCGCCAAGCGCGCGCTCAGCAACCAGCACCAGGTGCGCGTCGAGGTCGAGTC 980

Qy       1413 CGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGC 1472
              |  |   ||  ||       | ||   ||      |  | | |  |       ||
Db        981 CCTGTTCGACGGCGTCGACTTGTCGGAGCCGCTCTCCCGGGCGCGGTTCGAGGAGCTCAA 1040

Qy       1473 CAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAA 1532
              ||||  |  ||| |      |  | |||||    |  |      ||  |  || |
Db       1041 CAGCGACCTCTTCAAGAAGACGATGGTGCCGGTGAGGAAGGCCATGGCGGACGCCCGGCT 1100

Qy       1533 GGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCG 1592
              |  |  |  ||||||||      |  |  |  | ||          |  |  ||
Db       1101 GAGCAAGGGCGACATCGACGAGATCGTCCTCGTCGGCGGCAGCACCAGGATCCCCAAGGT 1160

Qy       1593 CGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGC 1652
              ||  |||| | ||  |     ||| |  |  |  | || |  |  |  |  |
Db       1161 GCAGCAGCTGCTCAAGGACTACTTCGGTGGCAAGGAGCCCAACCGCGGCGTCAACCCCGA 1220

Qy       1653 CGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAA 1712
              |||| |  |  | | ||| || |  |  | ||  |  ||| |  |  |  | | |
Db       1221 CGAGGCCGTGGCGTACGGCGCCGCCGTGCAGGCCAGCATCATCAGCGGCCACGTCGACGA 1280
```



RESULT 3
CL967755
LOCUS CL967755 2886 bp DNA linear GSS 21-SEP-2004
DEFINITION OsIFCC015718 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL967755
VERSION CL967755.1 GI:52390149
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 2886)
AUTHORS Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES Location/Qualifiers
source 1. .2886
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN

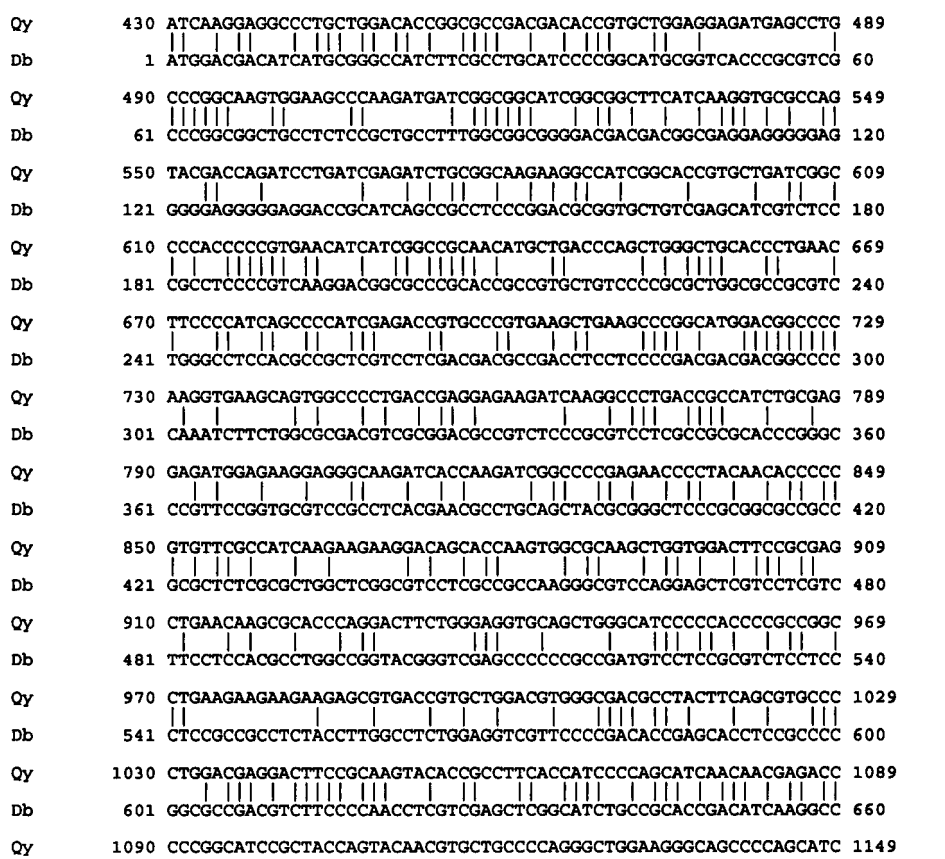
Query Match 3.9%; Score 97.2; DB 10; Length 2886;
Best Local Similarity 43.2%; Pred. No. 3.5e-07;
Matches 712; Conservative 0; Mismatches 903; Indels 33; Gaps 4;





```
RESULT 4
CL959255
LOCUS       CL959255               1509 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC002520 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL959255
VERSION     CL959255.1  GI:52373273
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1509)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .1509
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN
```

Query Match 3.7%; Score 91.4; DB 10; Length 1509;
Best Local Similarity 43.8%; Pred. No. 3.7e-06;
Matches 395; Conservative 0; Mismatches 506; Indels 0; Gaps 0;



```
Db        661 AAGGACCTCGACCGCCTGCTCCAGTGCAGCCCCGCGCTGGAGACGCTCGCGTTCGTCGTC 720

Qy       1150 TTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTG 1209

Db        721 AGCTACAACACGCAGCCCAACGTCCGCGTCCGCAGCCGCAGCCTCCGCTGCGTGCTCTTC 780

Qy       1210 ATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGC 1269

Db        781 TGGATGTCCATCGCGGAGGAGCTCGCCGTGGTCGTCGCCCCGCGGCTCGAGCGGCTCATC 840

Qy       1270 GCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAG 1329

Db        841 CTGTGGAACGGGTGCCCCTGCACCCGCCGTGTCCGCCGTGATTTCCGCACCAGGGTCAAG 900

Qy       1330 A 1330

Db        901 A 901
```

RESULT 5
AY103647
LOCUS AY103647 2598 bp mRNA linear HTC 18-FEB-2005
DEFINITION Zea mays PCO142084 mRNA sequence.
ACCESSION AY103647
VERSION AY103647.1 GI:21206725
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2598)
AUTHORS Gardiner,J., Schroeder,S., Polacco,M.L., Sanchez-Villeda,H., Fang,Z., Morgante,M., Landewe,T., Fengler,K., Useche,F., Hanafey,M., Tingey,S., Chou,H., Wing,R., Soderlund,C. and Coe,E.H. Jr.
TITLE Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization
JOURNAL Plant Physiol. 134 (4), 1317-1326 (2004)
PUBMED 15020742
REFERENCE 2 (bases 1 to 2598)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 3 (bases 1 to 2598)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES Location/Qualifiers
source 1. .2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeGDB:638378"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
ORIGIN

Query Match 3.7%; Score 90.8; DB 4; Length 2598;
Best Local Similarity 43.3%; Pred. No. 4.9e-06;
Matches 719; Conservative 0; Mismatches 909; Indels 33; Gaps 5;

```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db        351 GCAGCGAGTGGATCCTCGACTGCATCGCCCACGGCGGCGACATCTACGGCGTCACCACCG 410

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db        411 GCTTCGGCGGCACCTCCCACCGCCGCACCAAGGACGGGCCCGCGCTCCAGGTCGAGCTGC 470

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCG 349

Db        471 TCAGGCATCTCAACGCCGGAATCTTCGGCACCGGCAGCGACGGGCACACGCTGCCGTCGG 530

Qy        350 CCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGA 409

Db        531 AGGTCACCCGCGCGGCGATGCTGGTGCGCATCAACACCCTCCTCCAGGGCTACTCCGGCA 590

Qy        410 GCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCG 469

Db        591 TCCGCTTCGAGATCCTCGAGGCCATCACGAAGCTGCTCAACACCGGTGTCAGCCCCTGCC 650

Qy        470 TGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCG 529

Db        651 TGCCGCTCCGGGGCACCATCACCGCGTCGGGCGACCTGGTCCCGCTCTCCTACATCGCCG 710

Qy        530 GCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCA 589

Db        711 GCCTCATCACGGGCCGCCCCAACGCGCAGGCCGTCACCG---TCGACGGGAGGAAGGTGG 767

Qy        590 TCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCC 649

Db        768 ACGCCGCCGAGGCGTTCAAGATCGCCGGCATCGAGGGCGGCTTCTTCAAGCTCAACCCCA 827

Qy        650 AGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGA 709

Db        828 AGGAGGGCCTCGCCATCGTCAACGGCACGTCCGTGGGCTCCGCGCTCGCGGCCACCGTGA 887

Qy        710 AGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGG 769

Db        888 TGTACGACGCCAACGTCCTGGCCGTCCTGTCGGAGGTCCTGTCCGCCGTCTTCTGCGAGG 947

Qy        770 CCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCG 829

Db        948 TCATGAACGGCAAGCCCGAGTACACGGACCACCTGACCCACAAGCTGAAGCACCACCCGG 1007

Qy        830 AGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCA 889

Db       1008 GGTCCATCGAGGCCGCGGCCATCATGGAGCACATCCTGGATGGCAGCTCCTTCATGAAGC 1067

Qy        890 AGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGG 949

Db       1068 AGGCCAAGAAGTTGAACGAGCTGGACCCGC----------TGCTGAAGCCCAAGCAGG 1115

Qy        950 GCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCG 1009

Db       1116 ACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTGGGCCCCCAGATCGAGGTCATCCGCG 1175

Qy       1010 ACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCC 1069

Db       1176 CCGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTCATCGACG 1235

Qy       1070 CCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCT 1129

Db       1236 TCCACCGCGGCAAGGCGCTGCACGGCGGCAACTTCCAGGGCACCCCCATCGGCGTGTCCA 1295

Qy       1130 GGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCG 1189
```

```
Db       1296 TGGACAACGCCCGCCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCG 1355

Qy       1190 CCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGA-- 1247

Db       1356 AGCTCGTCAACGAGTTCTACAACAACGGGCTCACCTCCAACCTGGCCGGCAGCCGCAACC 1415

Qy       1248 ----CCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1303

Db       1416 CCAGCCTGGACTACGGCTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCCGAGC 1475

Qy       1304 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1363

Db       1476 TCCAGTACCTGGGCAACCCCATCACCAACCACGTGCAGAGCGCGGAGCAGCACAACCAGG 1535

Qy       1364 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1423

Db       1536 ACGTGAACTCCCTGGGCCTCGTCTCGGCCAGGAAGACCGCCGAGGCGATCGACATCCTGA 1595

Qy       1424 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1483

Db       1596 AGCTCATGTCGTCCACCTACATCGTGGCGCTGTGCCAGGCCGTGGACCTGCGCCACCTCG 1655

Qy       1484 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1543

Db       1656 AGGAGAACATCAAGGCGTCGGTGAAGAACACCGTGACCCAGGTGGCCAAGAAGGTGCTGA 1715

Qy       1544 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAA---------CCGCG 1594

Db       1716 CCATGAACCCCTCGGGCGAGCTCTCCAGCGCCCGCTTCAGCGAGAAGGAGCTGATCAGCG 1775

Qy       1595 AGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCG 1654

Db       1776 CCATCGACCGCGAGGCCGTGTTCACGTACGCGGAGGACGCGGCCAGCGGCAGCCTGCCGC 1835

Qy       1655 AGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGA 1714

Db       1836 TGATGCAGAAGCTGCGCGCCGTGCTGGTGGACCACGCCCTCAGCAGCG---GCGACGCGG 1892

Qy       1715 ACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGC 1774

Db       1893 AGCGGGAGCCCTCCGTGTTCTCCAAGATCACCAGGTTCGAGGAGGAGCTCCGCGCGGTGC 1952

Qy       1775 TGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTG 1815

Db       1953 TGCCCCAGGAGGTGGAGGCCGCCCGCGTGGCGGTCGCCGAG 1993
```

RESULT 6
CZ247380
LOCUS CZ247380 743 bp DNA linear GSS 11-FEB-2005
DEFINITION AIAA-aaf35a12.b1 Ancylostoma caninum whole genome shotgun library (AIAAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION CZ247380
VERSION CZ247380.1 GI:59632821
KEYWORDS GSS.
SOURCE Ancylostoma caninum (dog hookworm)
ORGANISM Ancylostoma caninum
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida; Ancylostomatoidea; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
REFERENCE 1 (bases 1 to 743)
AUTHORS Mitreva,M., McCarter,J.P., Pape,D., Ritter,E., Tsagareishvili,R., Ronko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Messina,D., Waterston,R.H., Clifton,S.W. and Wilson,R.
TITLE Genome Survey sequences from the parasitic nematode Ancylostoma caninum
JOURNAL Unpublished (2004)
COMMENT Contact: Mitreva M
Washington University in St. Louis
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@watson.wustl.edu
Genomic DNA provided by John Hawdon (mtmjmh@gwumc.edu) DNA sequenced by Washington University Genome Sequencing Center
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .743
/organism="Ancylostoma caninum"
/mol_type="genomic DNA"
/strain="Baltimore"
/db_xref="taxon:29170"
/dev_stage="Adult"
/lab_host="GS10"
/clone_lib="Ancylostoma caninum whole genome shotgun library (AIAAGSS 001)"
/note="Vector: pOTW13; Site_1: BstX1; Site_2: BstX1; Ancylostoma caninum genomic DNA was randomly sheared, end-repaired and size fractioned to enrich for 2-4 kb fragments. Genomic DNA was provided by John Hawdon (mtmjmh@gwumc.edu) at George Washington University. Sequencing by Washington University Genome Sequencing Center, St. Louis, MO."
ORIGIN

Query Match 3.6%; Score 89.4; DB 10; Length 743;
Best Local Similarity 46.3%; Pred. No. 8e-06;
Matches 294; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

```
Qy        751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAG 810

Db         63 AACGACAACAACAACGACAACGACAACAACGACTTCAACAACAACGACAACAACAACGAC 122

Qy        811 ATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAG 870

Db        123 AACAACAACAACGACGACAACAACGACAACAACGACAACAACAACGACAACAACAACAAC 182

Qy        871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930

Db        183 AACAACAACAACAACAACGACAACAACGACAACAACAACGACAACAACAACAACGACGAC 242

Qy        931 TTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTG 990

Db        243 GACAACAACGACAACAACGACAACAACAACAACAACAACAACGACAACAACAACGACAAC 302

Qy        991 ACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAG 1050

Db        303 AACAACAACGACGACGACAACAACGACAACAACGACAACGACAACAACAACAACAACAAC 362

Qy       1051 TACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTAC 1110

Db        363 AACGACAACAACAACGACAACAACAACAACGACGACAACAACGACAACGACAACGACAAC 422

Qy       1111 AACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170

Db        423 AACAACGACAGCAACAACGACAACGACAACAACAACAACAACGACAACAACGACAACAAC 482

Qy       1171 ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGAC 1230

Db        483 AACAACAACGACAACAACAACGACAACAACAACGACAACAACAACAACAACAACAACGAC 542

Qy       1231 CTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGC 1290

Db        543 AACGACAACGACAACGACTACGACAACAACAACGACAACAACAACATCGACGACTACAAC 602

Qy       1291 AAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCC 1350

Db        603 AACAACGACAACAACAACGACGACAACTACAACAACGACAACTACTACAACGACAACAAC 662

Qy       1351 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAA 1385

Db        663 TACAACGACAACGACAACAACAACGACGACTACTA 697
```

RESULT 7
CL961989

```
LOCUS       CL961989               1398 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC006991 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL961989
VERSION     CL961989.1  GI:52378720
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1398)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .1398
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.5%;  Score 86.6;  DB 10;  Length 1398;
  Best Local Similarity   45.9%;  Pred. No. 2.7e-05;
  Matches  508;  Conservative    0;  Mismatches  569;  Indels   30;  Gaps    5;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              |||||       | | | ||||||||   | |      |  | ||  ||||  ||  |   ||
Db        166 GTCGAGAGGCACGTCGGCGAGGCCGACATCGTCTTCGTCTCCGTGAACACCCCCACCAAG 225

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
                | ||  | ||   | |  ||||  | |    |   | | |||| |  |  | ||
Db        226 GCCCGCGGCCTCGGCGCCGGCAAGGCCGCCG---ACCTCACCTACTGGGAGAGCGCGGCG 282

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              |  ||   | |  |       |  ||  ||||   | |  |  ||||| |  |  |  |
Db        283 CGGATGATCGCCGCCGTGGCGACGTCCGACAAGGTCGTCGTCGAGAAGTCCACCGTGCCG 342

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              | | |       | ||| |  |    ||  ||  | |||  |  |||||| | | | | |
Db        343 GTCAAGACCGCGGAGGCCATCGAGAAGATCCTCGACCACAACGGCCGCGACGGCGTCGGC 402

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||      |    ||  ||||||| |  ||||   | |  |  |  |  ||    |
Db        403 TTCCAGATCCTCTCCAACCCGGAGTTCCTCGCCGAGGGCACCGCCATCCGCGACCTGCTC 462

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
                || |||          |||| |  || |  | || ||||| ||| ||| | ||||
Db        463 GCCCCCGA---------CCGCGTCCTCATCGGCGGCCGCGAGACCGCCGCGGGGCGCGCC 513

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              | |  || |    ||    | |   |    |  |  |||| || ||| |||    |
Db        514 GCCGTCCAGGCGCTCAAGGACGTGTACGCCCGATGGGTCCCCGAGGAGAGGATC---CTC 570

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |  | |  ||  |   |||  |||| ||    |||||  | ||| |  | ||||   ||
Db        571 ACCACCAACCTCTGGTCCGCCGAGCTGTCCAAGCTCGCCGCCAACGCGTTCCTGGCCCAG 630

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
```

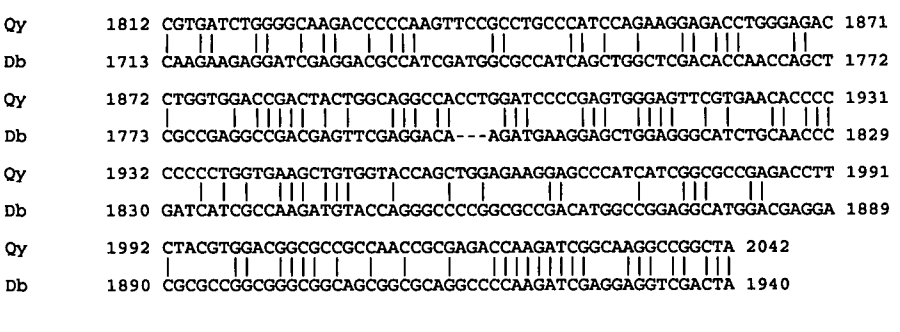
```
              | || |  | |||  |||    | | ||  | |   ||  | || |||||  |  | ||
Db        631 AGGATCTCGTCCGTCAACGCCATGTCCGCGCTCTGCGAGGCCACCGGCGCCGACGTCGCC 690

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGA---GATCTGCGGCAAGAAGGCCATCGGCACC 597
              | |  |   |||| | |  |  |  |     |||| ||| ||||     || || || |
Db        691 GAGGTCGCCTACGCCGTGGGCAAGGACTCCAGGATCGGCGCCAAGTTTCTCAACGCCAGC 750

Qy        598 GTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGC 657
              |||    ||||   | ||  |  | |    |  || |||| |||||
Db        751 GTGGGATTCGGTGGCTCCTGCTTCCAGAAAGACATCCTCAACCTGCTG------------ 798

Qy        658 TGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGC 717
              | || |    | | |      |  ||||   | | ||      |  |||||  |      |
Db        799 TACATCTGCGAGTGCAATGGCCTCCCCGAGGTGGCCAACTACTGGAAGCAGGTTATCAAG 858

Qy        718 ATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACC 777
              ||  |||  |  |||  ||  | |||                ||| | |  | ||  |  |||
Db        859 ATCAACGACTACCAGAAGAGCCGGTTCGTGAACCGCGTGGTGTCGTCCATGTTCAACACC 918

Qy        778 GCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCC 837
              |    | || || |||| |    |  ||||     |  |||||  | | |||   || ||
Db        919 GTGGCCGGCAAGAAGATCGCCGTGCTGGGCTTCGCCTTCAAGAAGGACACCGGCGACACC 978

Qy        838 TACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTG 897
                | || ||||   ||| | ||   |||  |   | |  | |   |  |||| ||
Db        979 AGGGAGACGCCCGCCATCGACGTCTGCAAGGGGCTGATCGGCGACAAGGCCAAGGTGAGC 1038

Qy        898 GACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCC 957
              || |  |   | |  | |  |||||  |   | ||  |    || |||    |
Db       1039 ATCTACGACCCCCAGGTGACGGAGGACCAGGTCCAGCGCGACCTCGCCATGAGCAAGTTC 1098

Qy        958 CACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTAC 1017
              |  |    || |  |    ||  ||| |   || | |   |  | |  |  | | | |
Db       1099 GATTGGGATCACCCGGTCCATCTCCAGCCGATGAGCCCCACGGCCATCAAGCAGGTGTCC 1158

Qy       1018 TTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATC 1077
              ||  |  |  |    ||||||    | |  |   |  || | ||  ||||| || |
Db       1159 GTCGCCTGGGACGCCTACGAGGCGGCCAGGGACGCCCACGGCGTCTGCATCCTCACCGAG 1218

Qy       1078 AACAACGAGACCCCCGGCATCCGCTAC 1104
                |||||  |   || ||  ||||
Db       1219 TGGGACGAGTTCAGGAGCCTCGACTAC 1245
```

```
RESULT 8
BM321451
LOCUS       BM321451                951 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
            3.3.1.1), mRNA sequence.
ACCESSION   BM321451
VERSION     BM321451.1  GI:18055857
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 951)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 951   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .951
```

```
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.5%;  Score 85.6;  DB 3;  Length 951;
  Best Local Similarity   46.0%;  Pred. No. 3.9e-05;
  Matches  323;  Conservative    0;  Mismatches  376;  Indels    3;  Gaps    1;

Qy       1734 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1793
             ||||||||| |||       ||||  | |    ||  |  | | || |
Db        219 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 278

Qy       1794 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1853
              |||||   ||   || ||  | |||  | |  ||    | ||  || || |     |
Db        279 CATCGCGCAGGCCGGCGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 338

Qy       1854 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1913
              | ||   |||||| || ||  |||   || |  ||  |||| | |   ||||  ||
Db        339 GGAGTGCACCTGGAAGGCCCTGTGCTTCGGCCCCTACCAGGGCCCTCAGATCATCGTCGA 398

Qy       1914 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1973
              ||   ||   || |  | |    || | |  | |||| |   |||  | ||||
Db        399 CGACGGCGGTGACGCGACTCTGATGATCCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 458

Qy       1974 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2033
              |    | ||   |    | | || |  | ||||  || | | | |  | ||    |
Db        459 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 518

Qy       2034 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2093
              || |   | | |   | |  | |||| |||||| |   |  || |          ||
Db        519 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 578

Qy       2094 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2153
              | || ||||   |      |  ||  ||||  || || | || ||| | ||| || ||
Db        579 CGAGGAGACGACGACTGGCGTNATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 638

Qy       2154 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2213
              ||  | ||| || |    ||| |       ||| | ||  | ||  |||||| | | |
Db        639 GTTCCCGGCCGTCAACGTCAACGAC---TCTNTCACCAAGAGCAAGTTCGACAACATCTA 695

Qy       2214 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2273
              ||   |    |  ||| ||      | |  |  ||   | ||  |    |
Db        696 CGGCTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGG 755

Qy       2274 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2333
              |  |||  ||| | |  |||| ||||| |    |  ||  | |  |    | |
Db        756 CAAGGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCG 815

Qy       2334 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
              |||    ||       |  || ||   | |||||   |  |  || | |  |    ||
Db        816 CGGCCAGGGCTGCCGCGTCATCGTGACGGAGATCGACCCCATCTGCGCGCTGCAGGCGTC 875

Qy       2394 CATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGA 2435
              |||| || | |  | ||   ||    ||   |  || ||||
Db        876 GATGGCCGGCTTCGAGGTCAACACGCTCGAGGCGGGGCTCGA 917


RESULT 9
CK159167/c
LOCUS       CK159167                869 bp    mRNA    linear   EST 05-DEC-2003
DEFINITION  FGAS040564 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA,
            mRNA sequence.
ACCESSION   CK159167
VERSION     CK159167.1  GI:38985053
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
            Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
            Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
            Penniket,C., Roach,J.L. and Sarhan,F.
  TITLE     Functional Genomics of Abiotic Stress In Wheat and Canola Crops
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas_ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [128,636].
            Plate: TaLt537  row: N  column: 23.
FEATURES             Location/Qualifiers
     source          1. .869
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Wheat line PI 178383"
                     /db_xref="taxon:4565"
                     /lab_host="DH5 alpha"
                     /clone_lib="Triticum aestivum FGAS: TaLt5"
                     /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
                     subtractive hybridization) cDNA library from genotype
                     PI178383 cold hardened at 2 C for 21 days and 49 days
                     (equal amount of cDNA pooled together before subtraction,
                     tester) and subtracted against genotype Norstar cold
                     hardened at 2 C for 1 day (24 H)(driver). Modified Smart
                     cDNA (Clontech)priming and non-directional cloning"
ORIGIN

  Query Match             3.4%;  Score 85;  DB 7;  Length 869;
  Best Local Similarity   45.4%;  Pred. No. 5e-05;
  Matches  304;  Conservative    0;  Mismatches  365;  Indels    0;  Gaps    0;

Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
              | |  |||   | ||||   |  ||||  | |      || ||||| |  |  || | ||
Db        827 ATGGCGGCGGCCCGCACAACGAGGATCACCAACCAGGACGAGAACAACCACAACCACCAC 768

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
                  | |     ||  |||   || |  |  || || |  || | | | |       |
Db        767 CACAACGACGACAACAGCGACACAACGACCACCAACAACACCAACAACAACAACAACAAC 708

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
                ||    ||   | ||   || |  ||| |  || ||    | |     | | |
Db        707 ACCAACAACAACAACAACAACAACAACAACAACGACAACCACCACCACAACCACAACAAC 648

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
              || | ||| | |   | |  || | | | | |   | ||  |   ||| | || |||
Db        647 AACAACAACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 588

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
              | |  |  | | |||  | |||||| |  |      || || ||| || ||  ||| || |
Db        587 AACAACAACAACAACAACAACAACAACAACAACAACGACAACAACAACAACAACAACAAC 528

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
              ||      |||          || | | |     |||||   || | |  ||  |    |
Db        527 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 468

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
                  |  || |  | ||  | || |   |  | ||  |   ||   ||
Db        467 AACAACAACAACAACAACAACAACAACGACGACGACGACAACGACGACGACGACGACGAC 408

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
              |||     ||||  | ||  || |     |  ||| |  ||  |  |||  ||| |  |
```

```
Db        407 GACAACAACGACAACAACGACAACAACAACAACGACAACAACAACAACAACAACAACAAC 348

Qy       1060 TTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119

Db        347 GACCACAACGACAACGACAACAACAACAACAACAACAACAACAACGACAACAACCACAAC 288

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179

Db        287 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 228

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTG 1239

Db        227 AACAACAACAACAACAACGACGACAACGACGTCAACGATAACAACGACAACAACGACAAC 168

Qy       1240 GGCAGCGAC 1248

Db        167 AACGGCAAC 159


RESULT 10
CL973991
LOCUS       CL973991               3069 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC025416 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL973991
VERSION     CL973991.1  GI:52402507
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 3069)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .3069
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.4%;  Score 84;  DB 10;  Length 3069;
  Best Local Similarity   42.8%;  Pred. No. 8.2e-05;
  Matches  536;  Conservative    0;  Mismatches  710;  Indels    6;  Gaps    2;

Qy         37 ACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAG 96

Db        625 AGCGGCATCGGCAAGACGACGCTGGCGAGCTACCTCTACCGGAGGAACAGCATCCGGAAG 684

Qy         97 TGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAG 156

Db        685 CACTTCGCCTGCACGGCGTGGGTCCACGTCCCCCGCAAGTTCCGCTTCGCCGACCTCCTC 744

Qy        157 GGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCC 216

Db        745 CACGACGTCATCCGCCAGGCCAACGCCTCCTCCGGCGACGACGAGCTCGGCGGCGAGGC- 803

Qy        217 AACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAG 276

Db        804 --CGCCGACGACGACGTGCGGCGCCGGGGTGGGATGGAGGCGGAGAGGCTCGTCAAGGCG 861

Qy        277 CAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGC 336

Db        862 CGCCTCGCCGCCGCGCTCCAGGGGAAGAGGTACCTCGTGGTGCTCGACGACGTCCGGAGC 921

Qy        337 AGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAG 396

Db        922 AGGGAGGAGTGGGCCTTCTTCCTCGCCGCGCTCCCCGGCGGCCTCGCCGGCAGCTGCGTG 981

Qy        397 CGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGC 456

Db        982 CTCGTCACGACGCAGGTCAAGATCAGCGAGGAGTCGGAGCAACCCGTGGCCGCCGGCGGC 1041

Qy        457 GCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATC 516

Db       1042 GGCGTCGTGCGCACCAGGGAGCTCGGGAAGCTGCGGCCGGAGCAGGCCGCCGAGCTGTTC 1101

Qy        517 GGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGC 576

Db       1102 CGCCGGAGGGTGTACGGCCACGGCGAGCCGGAGCCGAAGAAGATGGAGCAGCTCAAGAGC 1161

Qy        577 GGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGC 636

Db       1162 CTCGTGGACTCCATGACGAAGGGCTCCAACCTGCCGCTCAACATCGTCATGCTCGCCGGA 1221

Qy        637 AACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTG 696

Db       1222 CTCTTGAGGTCCAAGAAGGAGGATGAATGGGAGGCCGTGATCCATAGCCTCGATGATACG 1281

Qy        697 CCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAG 756

Db       1282 CCACCCCCGACATCGCCGGAGCAGACGCCGCCGGAGGCGGAGCCGAAGAAGGACAGCCTC 1341

Qy        757 GAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACC 816

Db       1342 GATGATAGCGAGACGGCGCCGACATCGACGGAGCAAACGAAGAAGACGAAGAAGAGGAAG 1401

Qy        817 AAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGC 876

Db       1402 CAGATGGCGACATCGATGGACAAGATACTGACGGTGTGCATGGACGACCTGCCGACGCAT 1461

Qy        877 ACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG 936

Db       1462 CTGAAGCCGTGCTTCCTCTACTTCGCCGGGTTCACCGCGCAGACGCCGATCTGCGCCGGG 1521

Qy        937 GAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTG 996

Db       1522 AAGCTGGTCCGGCTGTGGGTCGCCGAGGGGTTCATGCAGGCGAAGAACGGGCAGACGGTG 1581

Qy        997 CTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACC 1056

Db       1582 GAGGAGCACGGCGAGGAGTGCCTCAAGGAGCTCATCTCCCGGTGCCTCGTCCAGCTCGTC 1641

Qy       1057 GCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTG 1116

Db       1642 GAGACGGACGCCGGCGGCGGCAGGGTCGCCGCCGTCAGCATCCACCAAGCGGTGCTCGAC 1701

Qy       1117 CTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCA---GAGCAGCATGACCAAGATC 1173

Db       1702 TTCGTCCAGGCGGAGGCGCGCGACACCAACTTCCTCCACGTGCACAGCGGCGCGGCGGGC 1761

Qy       1174 CTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTG 1233

Db       1762 CTCTCCAACGGCGCCGCCCGCCGCCTCGCGCTCCGCAACACCTACGACACCGACCTCGCC 1821

Qy       1234 TACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGC 1285

Db       1822 GTCATGCTCGAGGCGCCCAAGCTCCACACCCTCCTCTGCGACATCCCGGAGC 1873
```

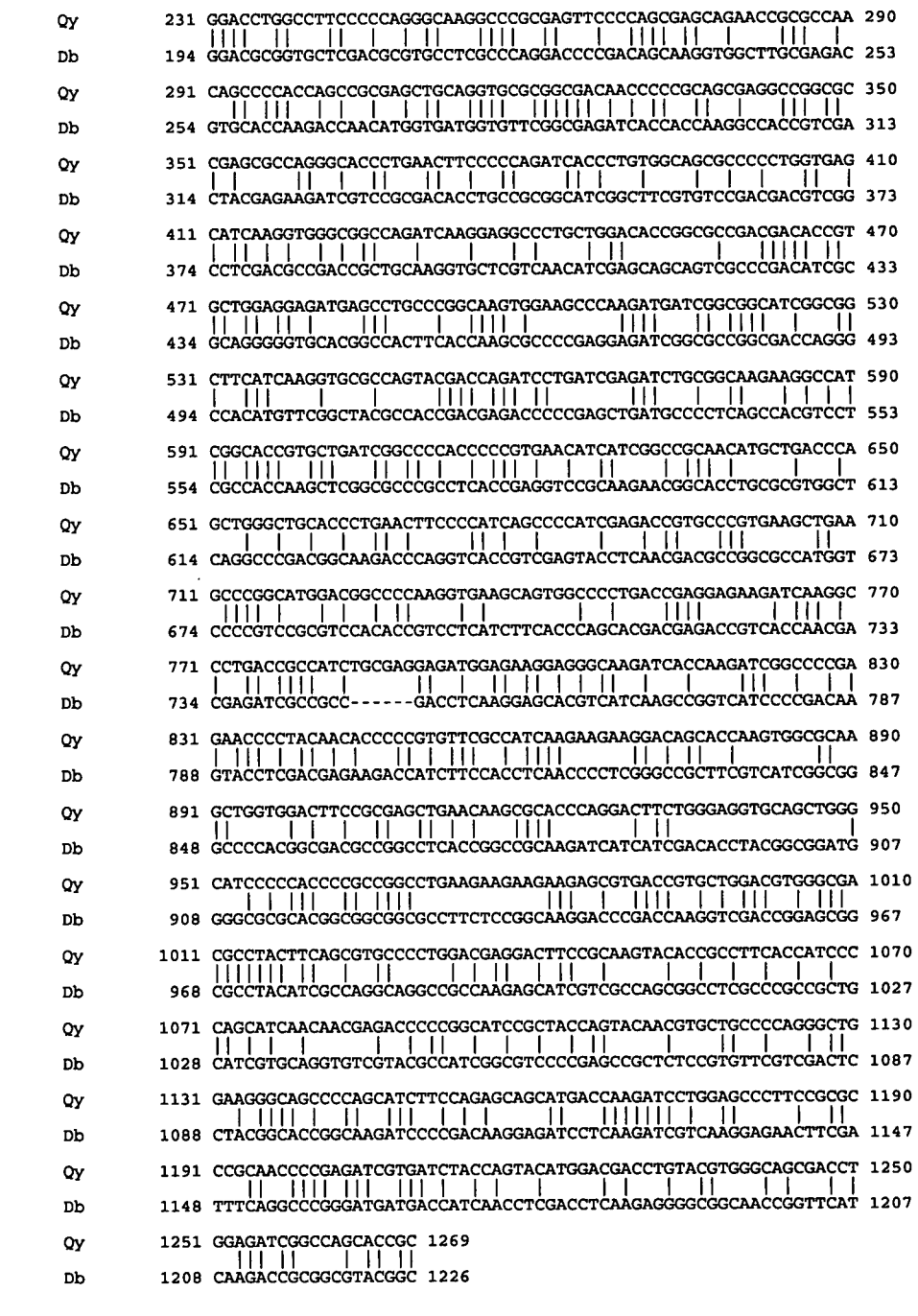
```
RESULT 11
CL971508
LOCUS       CL971508               1941 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC021485 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL971508
VERSION     CL971508.1  GI:52397596
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1941)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .1941
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.4%;  Score 83.8;  DB 10;  Length 1941;
  Best Local Similarity   42.9%;  Pred. No. 8.6e-05;
  Matches  769;  Conservative    0;  Mismatches  992;  Indels   30;  Gaps    6;

Qy        267 CCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGA 326
Db        165 CGCCGCCAAGAACCAGGTCGCCATGAACCCAACCAACACCGTCTTCGATGCGAAGCGGTT 224

Qy        327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCAC 386
Db        225 GATTGGTAGGAGGTTCTCCGACCCGTCCGTGCAGAGTGACATGAAGCTATGGCCGTTCAA 284

Qy        387 CCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCT 446
Db        285 GGTGGTGCCTGGCCCCGGCGACAAGCCGATGATCGTCGTCCAGTACAAGGGCGAGGAGAA 344

Qy        447 GGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCC 506
Db        345 GCAGTTCGCCGCGGAGGAGATCTCGTCCATGGTGCTGATCAAGATGAGGGAGATCGCGGA 404

Qy        507 CAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGAT 566
Db        405 GGCCTACCTCGGCTCGTCCATCAAGAACGCCGTGGTCACCGTCCCGGCCTACTTCAACGA 464

Qy        567 CGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACAT 626
Db        465 CTCGCAGAGGCAGGCCACCAAGGACGCCGGCGTCATCGCCGGGCTCAACGTGATGCGCAT 524

Qy        627 CATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCAT 686
Db        525 CATCAACGAGCCCACCGCCGCCGCCATCGCGTACGGCCTCGACAAGAAGGCGACGAGCAG 584

Qy        687 CGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCC 746
Db        585 CGGCGAGAAGAACGTGCTCATCTTTGACCTCGGCGGCGGCACGTTCGATGTGTCGCTGCT 644

Qy        747 CCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGG 806
Db        645 CACCATCGAGGAGGGCATCTTCGAGGTCAAGGCCACCGCCGGCGACACCCACCTCGGCGG 704

Qy        807 CAAGATCACCAAGATCGGCCCCGAGAACCCCTAC---AACACCCCCGTGTTCGCCATCAA 863
Db        705 CGAGGACTTCGACAACCGCATGGTGAACCACTTCGTGCAGGAGTTCAAGCGCAAGAACAA 764

Qy        864 GAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCAC 923
Db        765 GAAGGACATCAGCGGCAACCCGCGCGCGCTGCGGCGGCTCAGGACGGCGTGCGAGCGCGC 824

Qy        924 CCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAA 983
Db        825 CAAGCGCACGCTGTCGTCGACGGCGCAGACCACCATTGAGATCGACTCGCTGTACGAGGG 884

Qy        984 GAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTT 1043
Db        885 CATCGACTTCTACACGACGATCACCAGGGCGCGGTTCGAGGAGCTCAACATGGACCTGTT 944

Qy       1044 CCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTA 1103
Db        945 CCGCAAGTGCATGGAGCCCGTGGAGAAGTGCCTCCGCGACGCCAAGATGGACAAGAGCAG 1004

Qy       1104 CCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCAT 1163
Db       1005 CGTGCACGACGTCGTGCTC---------GTCGGCGGCTCCACCCGCATCCCCAAGGTGCA 1055

Qy       1164 GACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGTACAT 1223
Db       1056 GCAGCTGCTCCAGGACTTCTTCAACGGCAAGGAGCTCTGCAAGAGCATCAACCCCGACGA 1115

Qy       1224 GGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGA 1283
Db       1116 GGCCGTGGCGTACGGCGCCGCCGTCCAGGCCGCCATCCTGAGCGGCGAGGGCAACGAGAA 1175

Qy       1284 GCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGA 1343
Db       1176 GGTGC---AGGACCTGCTCCTGCTCGACGTCACCCCGCTGTCCCTCGGCCTCGAGACCGC 1232

Qy       1344 GCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC---GTGCAGCC 1400
Db       1233 CGGCGGCGTCATGACCGTGCTCATCCCGAGGAACACCACCATCCCGACCAAGAAGGAGCA 1292

Qy       1401 CATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAA 1460
Db       1293 GGTCTTCTCCACCTACTCCGACAACCAGCCCGGCGTCCTGATCCAGGTGTACGAGGGCGA 1352

Qy       1461 GCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA------- 1513
Db       1353 GCGCGCCAGGACCAAGGACAACAACCTCCTCGGCAAGTTCGAGCTCTCCGGCATCCCACC 1412

Qy       1514 --AGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGA 1571
Db       1413 GGCGCCGCGCGGCGTGCCCCAGATCAACGTCTGCTTCGACATCGACGCCAACGGCATCCT 1472

Qy       1572 GCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGA 1631
Db       1473 GAACGTGTCCGCGGAGGACAAGACGACGGGGCAGAAGAACAAGATCACCATCACCAACGA 1532

Qy       1632 CCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCA 1691
Db       1533 CAAGGGGCGGCTGAGCAAGGAGGAGATCGAGAAGATGGTGCAGGAGGCGGAGAAGTACAA 1592

Qy       1692 GATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGC 1751
Db       1593 GGCCGAGGACGAGGAGCACAAGAAGAAGGTGGACGCCAAGAACGCCCTCGAGAACTACGC 1652

Qy       1752 CCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCAT 1811
Db       1653 CTACAACATGCGCAACACCATCAAGGACGAGAAGATCGCCTCCAAGCTCGCCGCCGACGA 1712
```



```
Qy       1812 CGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGAC 1871
Db       1713 CAAGAAGAGGATCGAGGACGCCATCGATGGCGCCATCAGCTGGCTCGACACCAACCAGCT 1772

Qy       1872 CTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCC 1931
Db       1773 CGCCGAGGCCGACGAGTTCGAGGACA---AGATGAAGGAGCTGGAGGGCATCTGCAACCC 1829

Qy       1932 CCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTT 1991
Db       1830 GATCATCGCCAAGATGTACCAGGGCCCCGGCGCCGACATGGCCGGAGGCATGGACGAGGA 1889

Qy       1992 CTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTA 2042
Db       1890 CGCGCCGGCGGGCGGCAGCGGCGCAGGCCCCAAGATCGAGGAGGTCGACTA 1940
```

RESULT 12
CX099623
LOCUS CX099623 1781 bp mRNA linear EST 03-JUN-2005
DEFINITION RECm0260 A normalized whole-life-cycle cDNA library of rice Oryza sativa (indica cultivar-group) cDNA clone EI057O14, THREEB09, EI057O04, EI062D05, BI027I14, EI070G16, EI08 5', mRNA sequence.
ACCESSION CX099623
VERSION CX099623.1 GI:66912775
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 1781)
AUTHORS Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D., Han,B., Zhang,Q. and Wang,S.
TITLE Features of the expressed sequences revealed by a large-scale analysis of ESTs from a normalized cDNA library of the elite indica rice cultivar Minghui 63
JOURNAL Plant J. 42 (5), 772-780 (2005)
PUBMED 15918889
COMMENT Contact: Wang S
National Key Laboratory of Crop Genetic Improvement
Hauzhong Agricultural University
Wuhan 430070, China
Tel: 86-27-87282044
Fax: 86-27-87287092
Email: shipingwang@hotmail.com
Seq primer: T7.
FEATURES Location/Qualifiers
source 1. .1781
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/strain="indica"
/cultivar="Minghui 63"
/db_xref="taxon:39946"
/clone="EI057O14, THREEB09, EI057O04, EI062D05, BI027I14, EI070G16, EI08"
/tissue_type="whole plant"
/dev_stage="whole-life-cycle"
/lab_host="E. coli DH10B"
/clone_lib="A normalized whole-life-cycle cDNA library of rice"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The library is constructed based on the strategy of satuation hybridization with genomic DNA using rice cultivar Minghui 63. This library consists of cDNA from 15 directionally cloned cDNA libraries constructed with different tissues from 9 developmental stages."
ORIGIN

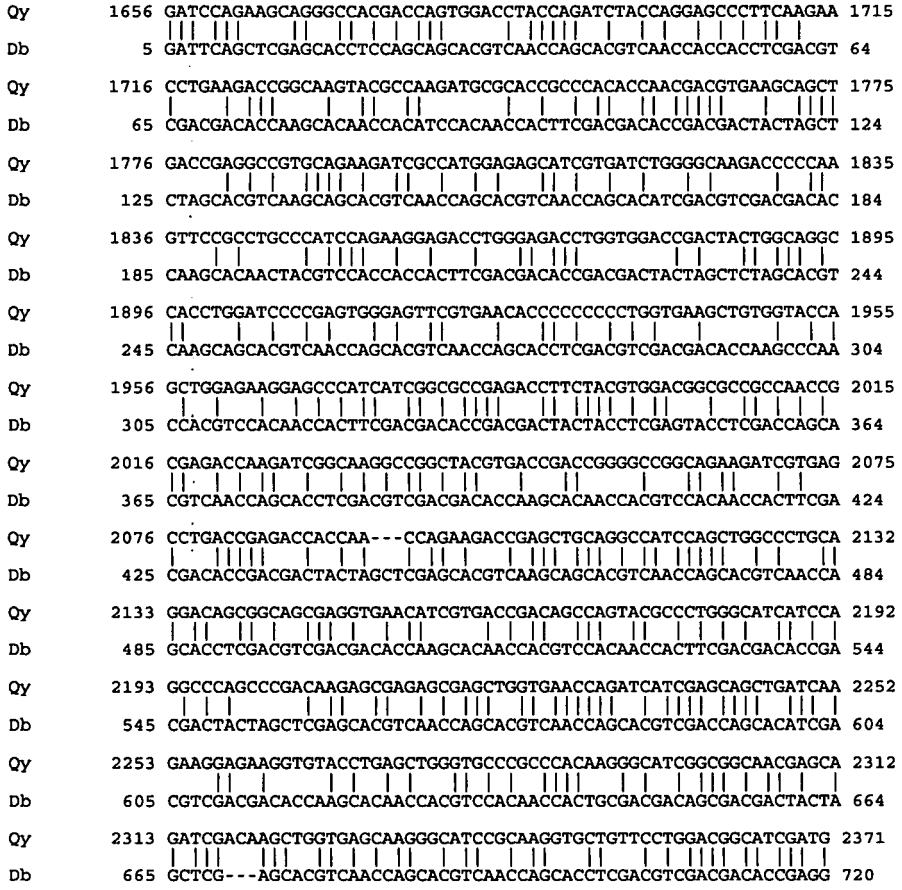
Query Match 3.3%; Score 82.6; DB 8; Length 1781;
Best Local Similarity 43.2%; Pred. No. 0.00014;
Matches 449; Conservative 0; Mismatches 584; Indels 6; Gaps 1;



```
Qy        231 GGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAA 290
Db        194 GGACGCGGTGCTCGACGCGTGCCTCGCCCAGGACCCCGACAGCAAGGTGGCTTGCGAGAC 253

Qy        291 CAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGC 350
Db        254 GTGCACCAAGACCAACATGGTGATGGTGTTCGGCGAGATCACCACCAAGGCCACCGTCGA 313

Qy        351 CGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAG 410
Db        314 CTACGAGAAGATCGTCCGCGACACCTGCCGCGGCATCGGCTTCGTGTCCGACGACGTCGG 373

Qy        411 CATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGT 470
Db        374 CCTCGACGCCGACCGCTGCAAGGTGCTCGTCAACATCGAGCAGCAGTCGCCCGACATCGC 433

Qy        471 GCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGG 530
Db        434 GCAGGGGGTGCACGGCCACTTCACCAAGCGCCCCGAGGAGATCGGCGCCGGCGACCAGGG 493

Qy        531 CTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCAT 590
Db        494 CCACATGTTCGGCTACGCCACCGACGAGACCCCCGAGCTGATGCCCCTCAGCCACGTCCT 553

Qy        591 CGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCA 650
Db        554 CGCCACCAAGCTCGGCGCCCGCCTCACCGAGGTCCGCAAGAACGGCACCTGCGCGTGGCT 613

Qy        651 GCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAA 710
Db        614 CAGGCCCGACGGCAAGACCCAGGTCACCGTCGAGTACCTCAACGACGCCGGCGCCATGGT 673

Qy        711 GCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGC 770
Db        674 CCCCGTCCGCGTCCACACCGTCCTCATCTTCACCCAGCACGACGAGACCGTCACCAACGA 733

Qy        771 CCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGA 830
Db        734 CGAGATCGCCGCC------GACCTCAAGGAGCACGTCATCAAGCCGGTCATCCCCGACAA 787

Qy        831 GAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAA 890
Db        788 GTACCTCGACGAGAAGACCATCTTCCACCTCAACCCCTCGGGCCGCTTCGTCATCGGCGG 847

Qy        891 GCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGG 950
Db        848 GCCCCACGGCGACGCCGGCCTCACCGGCCGCAAGATCATCATCGACACCTACGGCGGATG 907

Qy        951 CATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGA 1010
Db        908 GGGCGCGCACGGCGGCGGCGCCTTCTCCGGCAAGGACCCGACCAAGGTCGACCGGAGCGG 967

Qy       1011 CGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCC 1070
Db        968 CGCCTACATCGCCAGGCAGGCCGCCAAGAGCATCGTCGCCAGCGGCCTCGCCCGCCGCTG 1027

Qy       1071 CAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTG 1130
Db       1028 CATCGTGCAGGTGTCGTACGCCATCGGCGTCCCCGAGCCGCTCTCCGTGTTCGTCGACTC 1087

Qy       1131 GAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGC 1190
Db       1088 CTACGGCACCGGCAAGATCCCCGACAAGGAGATCCTCAAGATCGTCAAGGAGAACTTCGA 1147

Qy       1191 CCGCAACCCCGAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCT 1250
Db       1148 TTTCAGGCCCGGGATGATGACCATCAACCTCGACCTCAAGAGGGGCGGCAACCGGTTCAT 1207

Qy       1251 GGAGATCGGCCAGCACCGC 1269
Db       1208 CAAGACCGCGGCGTACGGC 1226
```

```
RESULT 13
CW922203/c
LOCUS       CW922203               1060 bp    DNA     linear   GSS 24-FEB-2005
DEFINITION  EDCAR29TR A. castellanii, 6-8 kb library from total genomic DNA
            Acanthamoeba castellanii genomic clone EDCAR29, genomic survey
            sequence.
ACCESSION   CW922203
VERSION     CW922203.1  GI:60228146
KEYWORDS    GSS.
SOURCE      Acanthamoeba castellanii
  ORGANISM  Acanthamoeba castellanii
            Eukaryota; Acanthamoebidae; Acanthamoeba.
REFERENCE   1  (bases 1 to 1060)
  AUTHORS   Anderson,I.J. and Loftus,B.J.
  TITLE     Gene discovery in the Acanthamoeba castellanii genome
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Iain Anderson
            The Institute for Genomic Research (TIGR; www.tigr.org)
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-795-7949
            Fax: 301-838-0208
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1060
                     /organism="Acanthamoeba castellanii"
                     /mol_type="genomic DNA"
                     /strain="Neff"
                     /db_xref="taxon:5755"
                     /clone="EDCAR29"
                     /clone_lib="A. castellanii, 6-8 kb library from total
                     genomic DNA"
                     /note="Vector: pHOS2"
ORIGIN

  Query Match             3.3%;  Score 80.8;  DB 10;  Length 1060;
  Best Local Similarity   43.9%;  Pred. No. 0.00029;
  Matches  392;  Conservative    0;  Mismatches  497;  Indels    3;  Gaps    1;

Qy        261 CGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCG 320
Db        955 CGTGCGCGGCGACGTGCCGACCAACCTCCAGGACCGCCGCGTCATCGCGCTCGACATGGG 896

Qy        321 CGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCA 380
Db        895 CGCGCTCGTGGCCGGCGCCAAGTACCGCGGCGAGTTCGAGGAGCGACTGAAGGCTGTGCT 836

Qy        381 GATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGC 440
Db        835 CAAGGAGGTCAAGGAGGCCGAGGGCAAGATCATCCTCTTCATCGACGAGATCCACCTCGT 776

Qy        441 CCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTG 500
Db        775 GCTGGGCGCCGGCAAGGCCGAGGGCGCCATGGACGCGGCCAACCTGCTCAAGCCCATGCT 716

Qy        501 GAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGAT 560
Db        715 CGCCCGCGGCGAGCTGCGCTGCATCGGCGCCACCACCCTGGCCGAGTACCGCAAGCACGT 656

Qy        561 CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGT 620
Db        655 GGAGAAGGACGCCGCGTTCGAGCGGCGCTTCCAGCCGGTCTACGTCGGCGAGCCCAGCGT 596

Qy        621 GAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAG 680
Db        595 CGAGGACACCGTGAGCATCCTGCGCGGGCTCAAGGAGAAGTACGAGGCCCACCACGGCGT 536

Qy        681 CCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCA 740
Db        535 GCGCATCATGGACTCGGCCTTGATCATGGCGTCCCAGCTCGCCGCCCGCTACATCACCAA 476

Qy        741 GTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA---TGGA 797
Db        475 CCGCTTCCTGCCCGACAAGGCCATCGATCTGATTGACGAAGCCTGCGCCAACACCCGCGT 416

Qy        798 GAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGC 857
Db        415 GCAGCTCGACTCGCAGCCCGAGGTGATCGACCAGCTCGAGCGCCGCCGTCTCAGGCTCGA 356

Qy        858 CATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAA 917
Db        355 GGTCGAGCTCACCGCCATGGAGAAGGAGTACAAGGAATCGAAGGACAAGGCCACCGAGCA 296

Qy        918 GCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAA 977
Db        295 GCGCATCGAGAAGATCAAGGAGGAGCTGTCGCGCATCAAGGAGGAGATCAGCCCGCTCAA 236

Qy        978 GAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGA 1037
Db        235 GGCGCGCTACGAGCAGGAGAAGGGCGCGCTCGACGAGATCCGCACGCTCAAGAAGCGACT 176

Qy       1038 GGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCAT 1097
Db        175 CGAGGACCTCAAGCTCAAGGCCGACGCCGCCGAGCGACGCAAGGACCTCGCCACCGCCGC 116

Qy       1098 CCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATC 1149
Db         115 CGATCTGCGCTACGGCGCCGTGCCCGAGCTCGAGAAGCGGATCGCCGCCCTC 64


RESULT 14
CZ216254
LOCUS       CZ216254                892 bp    DNA     linear   GSS 10-FEB-2005
DEFINITION  AIAA-aaf23g02.b1 Ancylostoma caninum whole genome shotgun library
            (AIAAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION   CZ216254
VERSION     CZ216254.1  GI:59229909
KEYWORDS    GSS.
SOURCE      Ancylostoma caninum (dog hookworm)
  ORGANISM  Ancylostoma caninum
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
            Ancylostomatoidea; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
REFERENCE   1  (bases 1 to 892)
  AUTHORS   Mitreva,M., McCarter,J.P., Pape,D., Ritter,E., Tsagareishvili,R.,
            Ronko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Messina,D.,
            Waterston,R.H., Clifton,S.W. and Wilson,R.
  TITLE     Genome Survey sequences from the parasitic nematode Ancylostoma
            caninum
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Mitreva M
            Washington University in St. Louis
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: nematode@watson.wustl.edu
            Genomic DNA provided by John Hawdon (mtmjmh@gwumc.edu) DNA
            sequenced by Washington University Genome Sequencing Center
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .892
                     /organism="Ancylostoma caninum"
                     /mol_type="genomic DNA"
                     /strain="Baltimore"
                     /db_xref="taxon:29170"
                     /dev_stage="Adult"
                     /lab_host="GS10"
                     /clone_lib="Ancylostoma caninum whole genome shotgun
                     library (AIAAGSS 001)"
                     /note="Vector: pOTW13; Site_1: BstX1; Site_2: BstX1;
                     Ancylostoma caninum genomic DNA was randomly sheared,
                     end-repaired and size fractioned to enrich for 2-4 kb
                     fragments. Genomic DNA was provided by John Hawdon
                     (mtmjmh@gwumc.edu) at George Washington University.
                     Sequencing by Washington University Genome Sequencing
                     Center, St. Louis, MO."
ORIGIN
```

Query Match 3.3%; Score 80.6; DB 10; Length 892;
Best Local Similarity 46.5%; Pred. No. 0.00031;
Matches 334; Conservative 0; Mismatches 379; Indels 6; Gaps 2;



RESULT 15
BM320864
LOCUS BM320864 1132 bp mRNA linear EST 03-JAN-2002
DEFINITION rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA sequence.
ACCESSION BM320864
VERSION BM320864.1 GI:18055270
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 1132)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 1132 Std Error: 0.00
POLYA=No.
FEATURES Location/Qualifiers
source 1. .1132
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"
ORIGIN

Query Match 3.2%; Score 80.2; DB 3; Length 1132;
Best Local Similarity 45.3%; Pred. No. 0.00037;
Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;

```
Qy        919 CGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              |  |   | |     | |||| |  ||||||  |||| || || ||
Db        850 AACCGCGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCC 897
```

Search completed: December 31, 2005, 02:31:20
Job time : 8727.85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 11955.3 Seconds
(without alignments)
11710.708 Million cell updates/sec

Title: US-09-610-313B-31
Perfect score: 2463
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2463

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2463	100.0	2463	6	AX455915	AX455915 Sequence
2	2442.2	99.2	2469	6	AX455914	AX455914 Sequence
3	2436.2	98.9	2457	6	AX455916	AX455916 Sequence
4	2046	83.1	2306	6	BD263704	BD263704 Improved
5	2046	83.1	2306	6	CQ870574	CQ870574 Sequence
6	2046	83.1	2306	6	AR373387	AR373387 Sequence
7	2027.6	82.3	9166	6	AX427930	AX427930 Sequence
8	2025.2	82.2	2312	6	BD263706	BD263706 Improved
9	2025.2	82.2	2312	6	CQ870576	CQ870576 Sequence
10	2025.2	82.2	2312	6	AR373389	AR373389 Sequence
11	2019.2	82.0	2300	6	BD263705	BD263705 Improved
12	2019.2	82.0	2300	6	CQ870575	CQ870575 Sequence
13	2019.2	82.0	2300	6	AR373388	AR373388 Sequence
14	2019.2	82.0	9788	6	AX427936	AX427936 Sequence
15	2005.2	81.4	9169	6	AX427931	AX427931 Sequence
16	2001.8	81.3	9194	6	AX427926	AX427926 Sequence
17	2000.2	81.2	9194	6	AX427925	AX427925 Sequence
18	2000.2	81.2	12411	6	AX427927	AX427927 Sequence
19	1996.8	81.1	9785	6	AX427938	AX427938 Sequence
20	1993.4	80.9	9167	6	AX427933	AX427933 Sequence
21	1993.4	80.9	9170	6	AX427928	AX427928 Sequence
22	1993.4	80.9	9782	6	AX427935	AX427935 Sequence
23	1993.4	80.9	9783	6	AX427934	AX427934 Sequence
24	1993.4	80.9	9790	6	CS078943	CS078943 Sequence
25	1993.4	80.9	9792	6	AX427932	AX427932 Sequence
c 26	1993.4	80.9	36066	6	CS078956	CS078956 Sequence
27	1991.8	80.9	9189	6	AX427921	AX427921 Sequence
28	1988.8	80.7	4353	11	AF287353	AF287353 Synthetic
29	1972.2	80.1	3015	6	AX455946	AX455946 Sequence
30	1955.6	79.4	3009	6	AX455987	AX455987 Sequence
31	1942.6	78.9	4319	6	BD263639	BD263639 Improved
32	1942.6	78.9	4319	6	CQ870498	CQ870498 Sequence
33	1942.6	78.9	4319	6	AR373322	AR373322 Sequence
34	1932	78.4	4352	11	AF287352	AF287352 Synthetic
35	1932	78.4	4352	11	AY789509	AY789509 Synthetic
36	1886.6	76.6	2577	6	AX457088	AX457088 Sequence
37	1879	76.3	4343	6	AX306428	AX306428 Sequence
38	1878.8	76.3	2305	6	BD263702	BD263702 Improved
39	1878.8	76.3	2305	6	CQ870572	CQ870572 Sequence
40	1878.8	76.3	2305	6	AR373385	AR373385 Sequence
41	1875	76.1	4341	6	AX306429	AX306429 Sequence
42	1865.4	75.7	1977	6	AX455954	AX455954 Sequence
43	1857.8	75.4	1965	6	AX455952	AX455952 Sequence
44	1857.2	75.4	7344	6	CS078948	CS078948 Sequence
45	1852.4	75.2	7897	6	AX427923	AX427923 Sequence

ALIGNMENTS

RESULT 1
AX455915
LOCUS AX455915 2463 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 31 from Patent WO0204493.
ACCESSION AX455915
VERSION AX455915.1 GI:21714900
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof
JOURNAL Patent: WO 0204493-A 31 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES Location/Qualifiers
source 1. .2463
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YM"
ORIGIN

Query Match 100.0%; Score 2463; DB 6; Length 2463;
Best Local Similarity 100.0%; Pred. No. 2e-202;
Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
```

```
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300
Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
```

```
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
```

```
Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2461 TTC 2463
             |||
Db       2461 TTC 2463


RESULT 2
AX455914
LOCUS       AX455914               2469 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 30 from Patent WO0204493.
ACCESSION   AX455914
VERSION     AX455914.1  GI:21714899
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 30 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1. .2469
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975(+)"
ORIGIN

  Query Match           99.2%;  Score 2442.2;  DB 6;  Length 2469;
  Best Local Similarity   99.6%;  Pred. No. 1.2e-200;
  Matches 2460;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              ||||||||||||||||      || |  ||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
```

```
Qy       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1915 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2455 GGTGAATTC 2463
             |||||||||
Db       2461 GGTGAATTC 2469
```

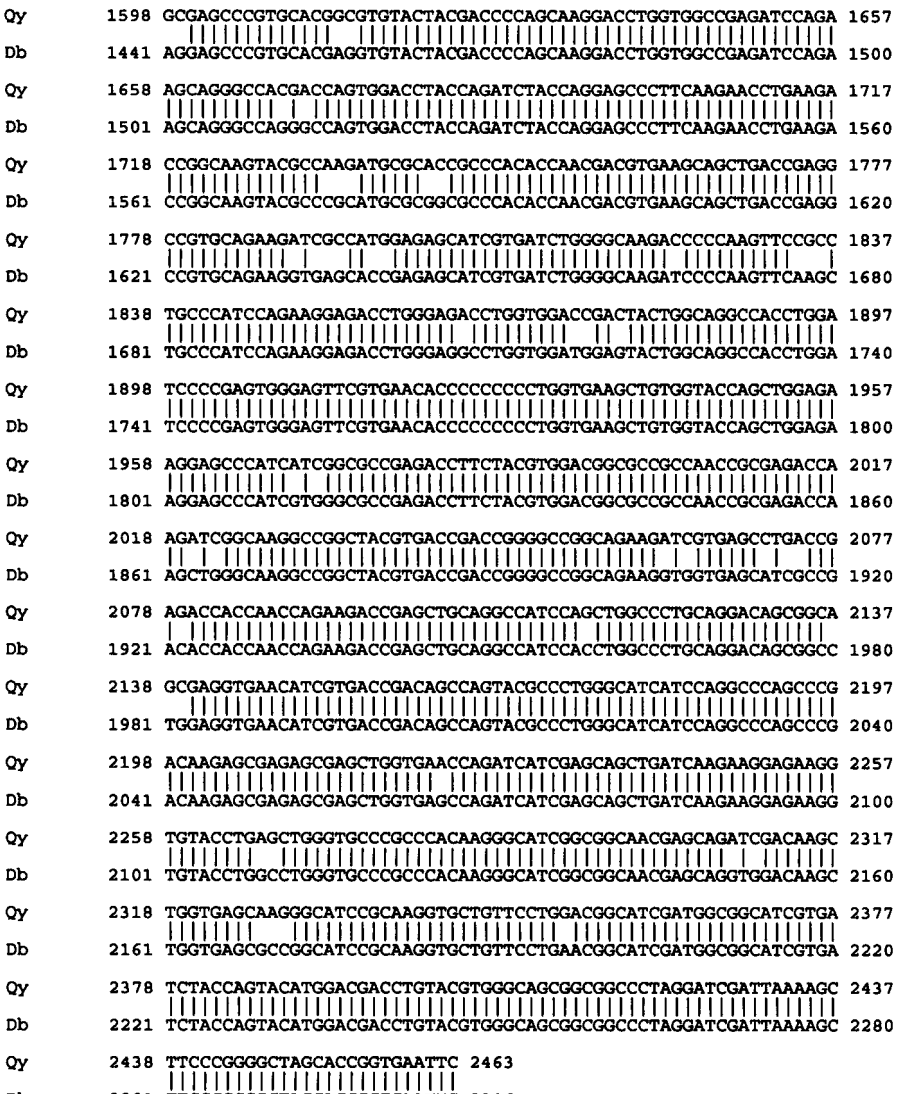
```
RESULT 3
AX455916
LOCUS       AX455916               2457 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 32 from Patent WO0204493.
ACCESSION   AX455916
VERSION     AX455916.1  GI:21714901
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 32 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1. .2457
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YMWM"
ORIGIN

  Query Match             98.9%;  Score 2436.2;  DB 6;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 4e-200;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
```

```
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Db       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Db       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Db       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Db       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Db       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Db       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Db       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Db       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Db       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Db       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Db       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Db       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Db       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Db       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Db       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Db       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Db       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Db       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454

Qy       2461 TTC 2463

Db       2455 TTC 2457
```

```
RESULT 4
BD263704
LOCUS       BD263704               2306 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263704
VERSION     BD263704.1  GI:33073472
KEYWORDS    JP 2002533124-A/71.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 71 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/71
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH   Key
                       Location/Qualifiers
            FT   source          1. .2306
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
```

```
     source          1. .2306
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             83.1%;  Score 2046;  DB 6;  Length 2306;
  Best Local Similarity   93.6%;  Pred. No. 1e-166;
  Matches 2159;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              |||||||||||||||  |||||||||||||||||||||  |||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              |||||||||| |||||||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      ||||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
              |||||||||||| ||||| | | ||||||||| |||||||||| ||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
              |||| ||||| |||||||||||||||| ||||||||||||||    | ||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
              || | |||||||||||||||||||| | ||||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
              || |||||||||| | |||||||||||||||||||||||||||||||||||||||| ||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
              |||||||||||||||| || |||||||||||||||| ||||||||||||||||||||||
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
              ||||||||||||||||  |  ||||||  |||||||||||||||||||||||||||| ||
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
              |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
              |||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||
Db         961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
              ||||||||||   |  |||||||| ||||||||||||||||||||||||||||||||||
Db        1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
              |||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||||
Db        1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
              |||||||||||||||||||||||||||||||||||||||||||  |||||||||||||| |
Db        1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
              || |||||||||||||  |||||||||||||||||||||| ||||||||||||||||||
Db        1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy       1538 ACATCGTGCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
              |  |  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
              |||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||
Db        1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
              |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db        1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
              |||||||||||||||   ||||||  ||||||||||||||||||||||||||||||||||
Db        1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
              |||||||||||  |   ||  |||||||||||||||||||||||||||||||||||||||
Db        1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
              || | |||||||||||||||||||||||||||||||||||| | |||||| |  |||
Db        1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
              | |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040
```

```
Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db       2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db       2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db       2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db       2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db       2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306


RESULT 5
CQ870574
LOCUS       CQ870574                2306 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 82 from Patent EP1433851.
ACCESSION   CQ870574
VERSION     CQ870574.1  GI:52000090
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 82 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.ROpt.YM"
ORIGIN

  Query Match             83.1%;  Score 2046;  DB 6;  Length 2306;
  Best Local Similarity   93.6%;  Pred. No. 1e-166;
  Matches 2159;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
Db       1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db       1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
Db       1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
Db       1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440
```



```
Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
Db       1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db       1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
Db       1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
Db       1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db       1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
Db       1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db       1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db       1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db       1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db       1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db       2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db       2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db       2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db       2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db       2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306
```

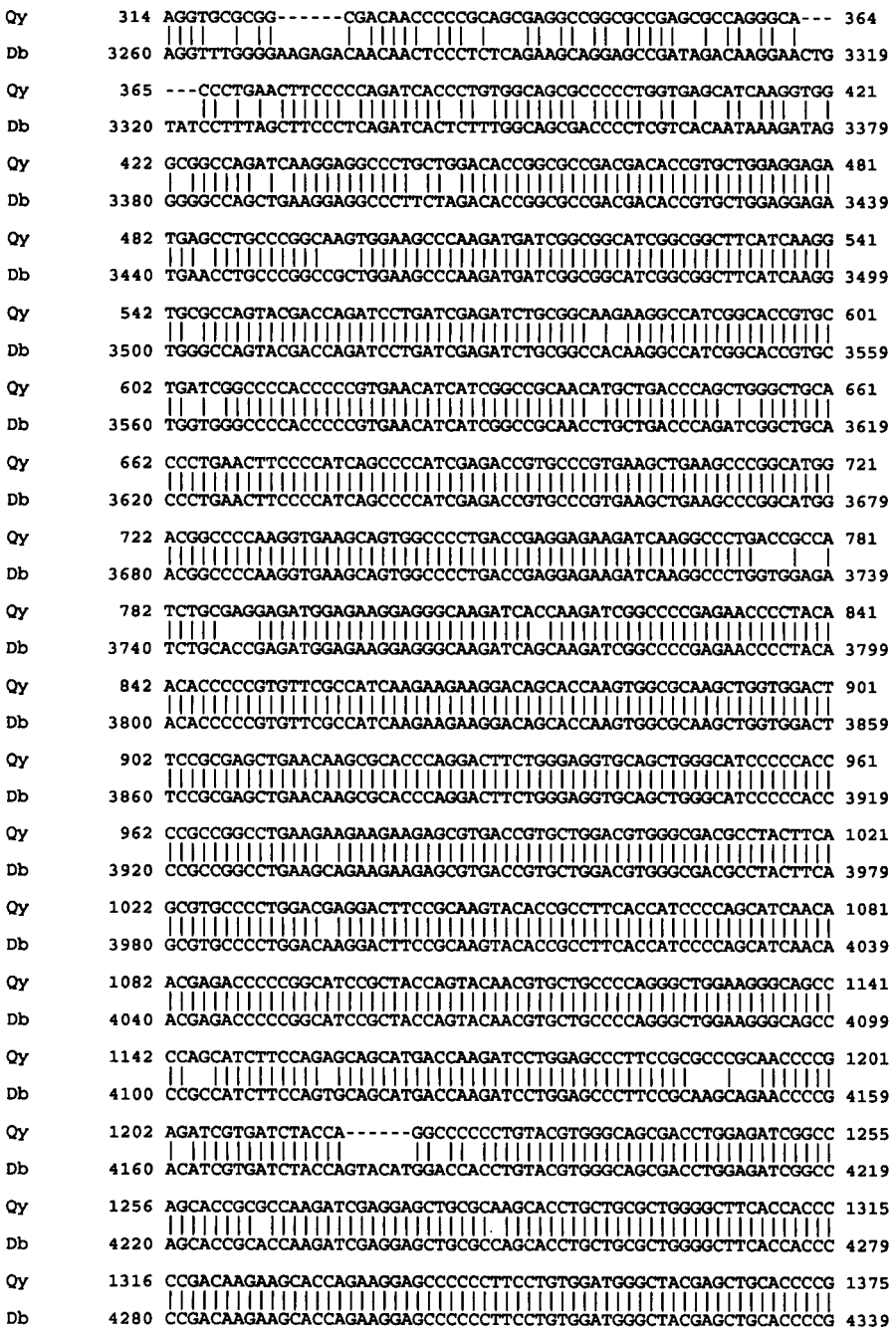
```
RESULT 6
AR373387
LOCUS       AR373387               2306 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 82 from patent US 6602705.
ACCESSION   AR373387
VERSION     AR373387.1  GI:40075490
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 82 05-AUG-2003;
            Chiron Corporation; Emeryville, CA
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

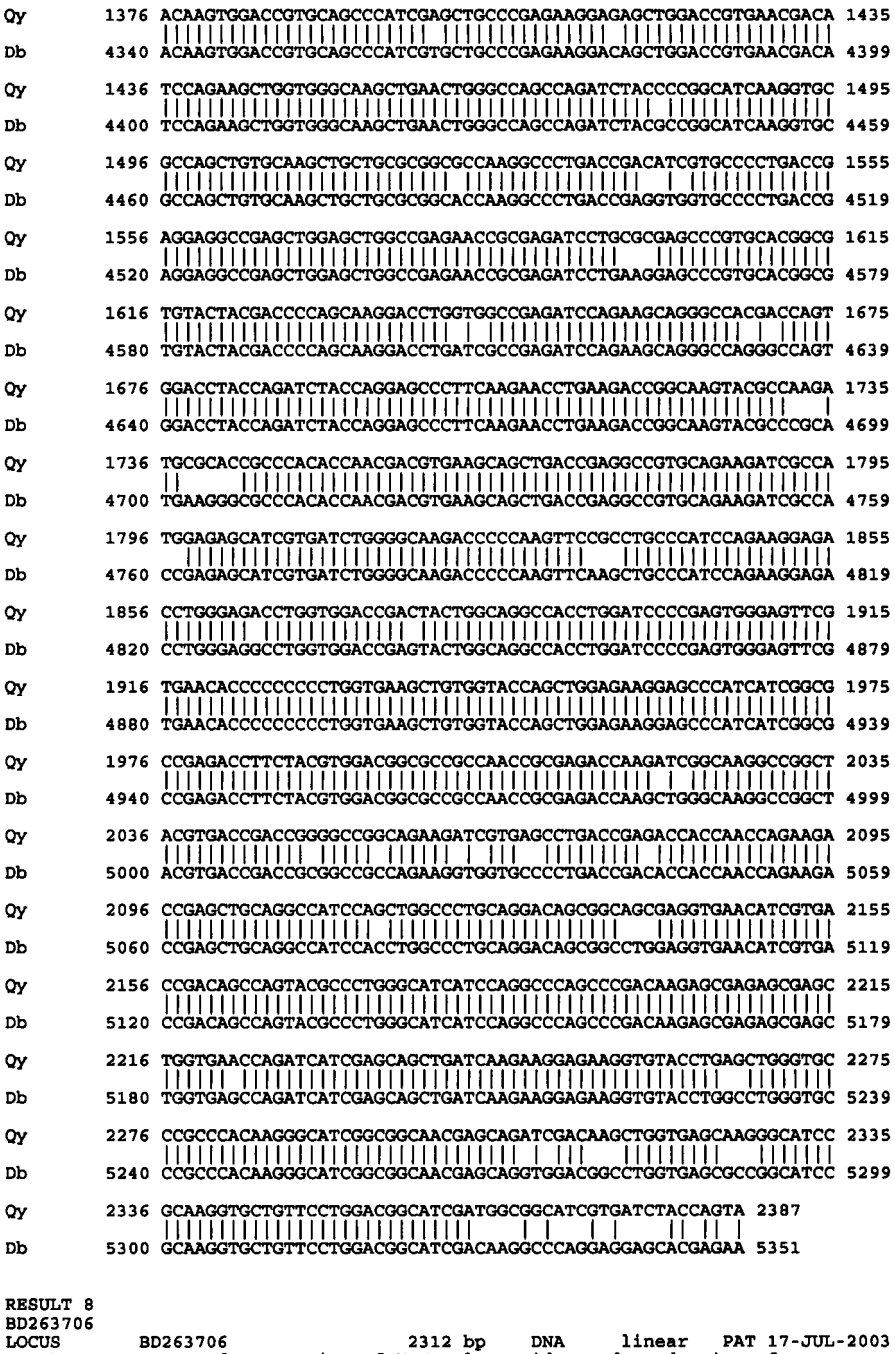
Query Match 83.1%; Score 2046; DB 6; Length 2306;
Best Local Similarity 93.6%; Pred. No. 1e-166;
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
```

```
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297

Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357

Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417

Db       1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477

Db       1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537

Db       1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597

Db       1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657

Db       1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717

Db       1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777

Db       1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837

Db       1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897

Db       1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957

Db       1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017

Db       1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077

Db       1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137

Db       1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197

Db       1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257

Db       2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317

Db       2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377

Db       2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437

Db       2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463

Db       2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306
```

```
RESULT 7
AX427930
LOCUS       AX427930               9166 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 168 from Patent WO0232943.
ACCESSION   AX427930
VERSION     AX427930.1  GI:21538017
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 168 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9166
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match           82.3%;  Score 2027.6;  DB 6;  Length 9166;
  Best Local Similarity   91.7%;  Pred. No. 2.4e-165;
  Matches 2194;  Conservative    0;  Mismatches  179;  Indels   19;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133

Db       3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193

Db       3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253

Db       3141 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3199

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313

Db       3200 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3259
```





RESULT 8
BD263706
LOCUS BD263706 2312 bp DNA linear PAT 17-JUL-2003
DEFINITION Improved expression of HIV polypeptides and production of

```
          virus-like particles.
ACCESSION   BD263706
VERSION     BD263706.1  GI:33073474
KEYWORDS    JP 2002533124-A/73.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 73 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/73
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH   Key
                 Location/Qualifiers
            FT   source          1. .2312
            FT                   /organism='Artificial Sequence'.
FEATURES        Location/Qualifiers
     source     1. .2312
                /organism="synthetic construct"
                /mol_type="genomic DNA"
                /db_xref="taxon:32630"
ORIGIN

  Query Match             82.2%;  Score 2025.2;  DB 6;  Length 2312;
  Best Local Similarity   93.3%;  Pred. No. 6.1e-165;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1351
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1411
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1412 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1471
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1472 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1531
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1532 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1591
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1592 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1651
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1652 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1711
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1712 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1771
```

```
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1772 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1831

Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1832 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1891

Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1892 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1951

Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1952 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2011

Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2012 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2071

Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2072 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2131

Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2132 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2191

Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2192 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2251

Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2252 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2311

Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2312 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2371

Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2372 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2431

Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2432 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2463

Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312


RESULT 9
CQ870576
LOCUS       CQ870576                2312 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 84 from Patent EP1433851.
ACCESSION   CQ870576
VERSION     CQ870576.1  GI:52000092
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 84 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt(+)"
ORIGIN

  Query Match             82.2%;  Score 2025.2;  DB 6;  Length 2312;
  Best Local Similarity   93.3%;  Pred. No. 6.1e-165;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343

Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637

Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697

Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757

Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817

Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877

Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997

Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
```

```
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291

Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1351

Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1411

Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1412 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1471

Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1472 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1531

Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1532 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1591

Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1592 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1651

Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1652 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1711

Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1712 TGAAGACCGGCAAGTACGCCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1771

Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1772 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1831

Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1832 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1891

Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1892 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1951

Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1952 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2011

Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2012 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2071

Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2072 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2131

Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2132 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2191

Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2192 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2251

Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2252 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2311

Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2312 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2371

Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2372 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2431

Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2432 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2463

Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 10
AR373389
LOCUS       AR373389               2312 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 84 from patent US 6602705.
ACCESSION   AR373389
VERSION     AR373389.1  GI:40075492
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 84 05-AUG-2003;
            Chiron Corporation; Emeryville, CA
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             82.2%;  Score 2025.2;  DB 6;  Length 2312;
  Best Local Similarity   93.3%;  Pred. No. 6.1e-165;
  Matches 2156;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343

Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420
```

```
Qy       578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db       421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy       638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db       481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy       698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db       541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy       758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db       601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy       818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db       661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy       878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db       721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy       938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db       781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy       998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db       841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy      1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db       901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy      1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db       961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy      1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
Db      1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy      1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
Db      1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy      1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1351
Db      1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy      1352 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1411
Db      1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy      1412 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1471
Db      1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy      1472 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1531
Db      1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy      1532 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1591
Db      1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy      1592 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1651
Db      1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy      1652 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1711

Db      1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy      1712 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1771
Db      1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy      1772 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1831
Db      1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy      1832 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1891
Db      1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy      1892 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1951
Db      1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy      1952 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2011
Db      1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy      2012 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2071
Db      1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy      2072 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2131
Db      1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy      2132 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2191
Db      1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy      2192 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2251
Db      2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy      2252 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2311
Db      2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy      2312 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2371
Db      2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy      2372 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2431
Db      2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy      2432 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2463
Db      2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

RESULT 11
BD263705

LOCUS	BD263705 2300 bp DNA linear PAT 17-JUL-2003
DEFINITION	Improved expression of HIV polypeptides and production of virus-like particles.
ACCESSION	BD263705
VERSION	BD263705.1 GI:33073473
KEYWORDS	JP 2002533124-A/72.
SOURCE	synthetic construct
ORGANISM	synthetic construct other sequences; artificial sequences.
REFERENCE	1 (bases 1 to 2300)
AUTHORS	Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.
TITLE	Improved expression of HIV polypeptides and production of virus-like particles
JOURNAL	Patent: JP 2002533124-A 72 08-OCT-2002;

```
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/72
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.ROpt.YMWM FH
            Key             Location/Qualifiers
            FT   source          1. .2300
            FT                   /organism='Artificial Sequence'.
FEATURES              Location/Qualifiers
     source           1. .2300
                      /organism="synthetic construct"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:32630"
ORIGIN

  Query Match             82.0%;  Score 2019.2;  DB 6;  Length 2300;
  Best Local Similarity   93.2%;  Pred. No. 2e-164;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              ||||||||||||||||  |||||||||||||||||||||  |||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              |||||||||| |||||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      |||||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
              |||||||||| ||||| | | ||||||||| |||||||||| ||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
              |||| ||||| |||||||||||||||| ||||||||||||||||    | ||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
              || | |||||||||||||||||||| | ||||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
              || |||||||||| | ||||||||||||||||||||||||||||||||||||| ||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
              |||||||||||||||| || |||||||||||||||| ||||||||||||||||||||||
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
              ||||||||||||||||    | |||||||   |||||||||||||||||||||||||| ||
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
              |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
              ||||||||||    | |||||||| ||||||||||||||||||||||||||||||||||
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
              |||||||||||||||||||||||||||| |||||||||||||||||||| ||||||||||
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1198

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
                 ||||||||||||||||||||||||||||||||||||  |||||||||||||||| |
Db       1199 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1254

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1314

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
              || |||||||||||||   |||||||||||||||||||||||| ||||||||||||||||
Db       1315 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1374

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
              |  |  | |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1434

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
              |||||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db       1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1494

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
              |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1554

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
              |||||||||||||||   |||||||  |||||||||||||||||||||||||||||||||
Db       1555 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1614

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
              |||||||||| |    || |||||||||||||||||||||||| ||||||||||    |
Db       1615 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
              |||||||||||||||||||||||||| ||||||||  || ||||||||||||||||||||
Db       1675 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1734

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1794
```

```
Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
Db       1795 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1854

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
Db       1855 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1914

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db       1915 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
Db       1975 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2034

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
Db       2035 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2094

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db       2095 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2154

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
Db       2155 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2214

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
Db       2215 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2274

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db       2275 TTCCCGGGGCTAGCACCGGTGAATTC 2300
```

```
RESULT 12
CQ870575
LOCUS       CQ870575               2300 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 83 from Patent EP1433851.
ACCESSION   CQ870575
VERSION     CQ870575.1  GI:52000091
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 83 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2300
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt.YMWM"
ORIGIN

  Query Match             82.0%;  Score 2019.2;  DB 6;  Length 2300;
  Best Local Similarity   93.2%;  Pred. No. 2e-164;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
```

```
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCCTTCCTGTGGATGG 1357
```

```
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1198

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417
                ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||| |
Db       1199 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1254

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1314

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537
                || ||||||||||||||    |||||||||||||||||||||||| ||||||||||||||
Db       1315 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1374

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597
                |   |  | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1434

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657
                  ||||||||||||  |||||||||||||||||||||||||||||||||||||||||||
Db       1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1494

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
                |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1554

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777
                |||||||||||||   ||||||  ||||||||||||||||||||||||||||||||||||
Db       1555 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1614

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837
                |||||||||||  |   ||  |||||||||||||||||||||||| |||||||||   |
Db       1615 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
                |||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db       1675 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1734

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1794

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017
                |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1854

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077
                || | ||||||||||||||||||||||||||||||||||||||||| | |||||| |  |||
Db       1855 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1914

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
                | |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1915 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2034

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2094

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
                |||||||  |||||||||||||||||||||||||||||||||||||| | |||||||
Db       2095 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2154

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377
                |||||||   |||||||||||||||||||||||| |||||||||||||||||||||||
Db       2155 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2214

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2274

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
                |||||||||||||||||||||||||
Db       2275 TTCCCGGGGCTAGCACCGGTGAATTC 2300
```

RESULT 13
AR373388
LOCUS AR373388 2300 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 83 from patent US 6602705.
ACCESSION AR373388
VERSION AR373388.1 GI:40075491
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2300)
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: US 6602705-A 83 05-AUG-2003;
Chiron Corporation; Emeryville, CA
FEATURES Location/Qualifiers
source 1. .2300
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

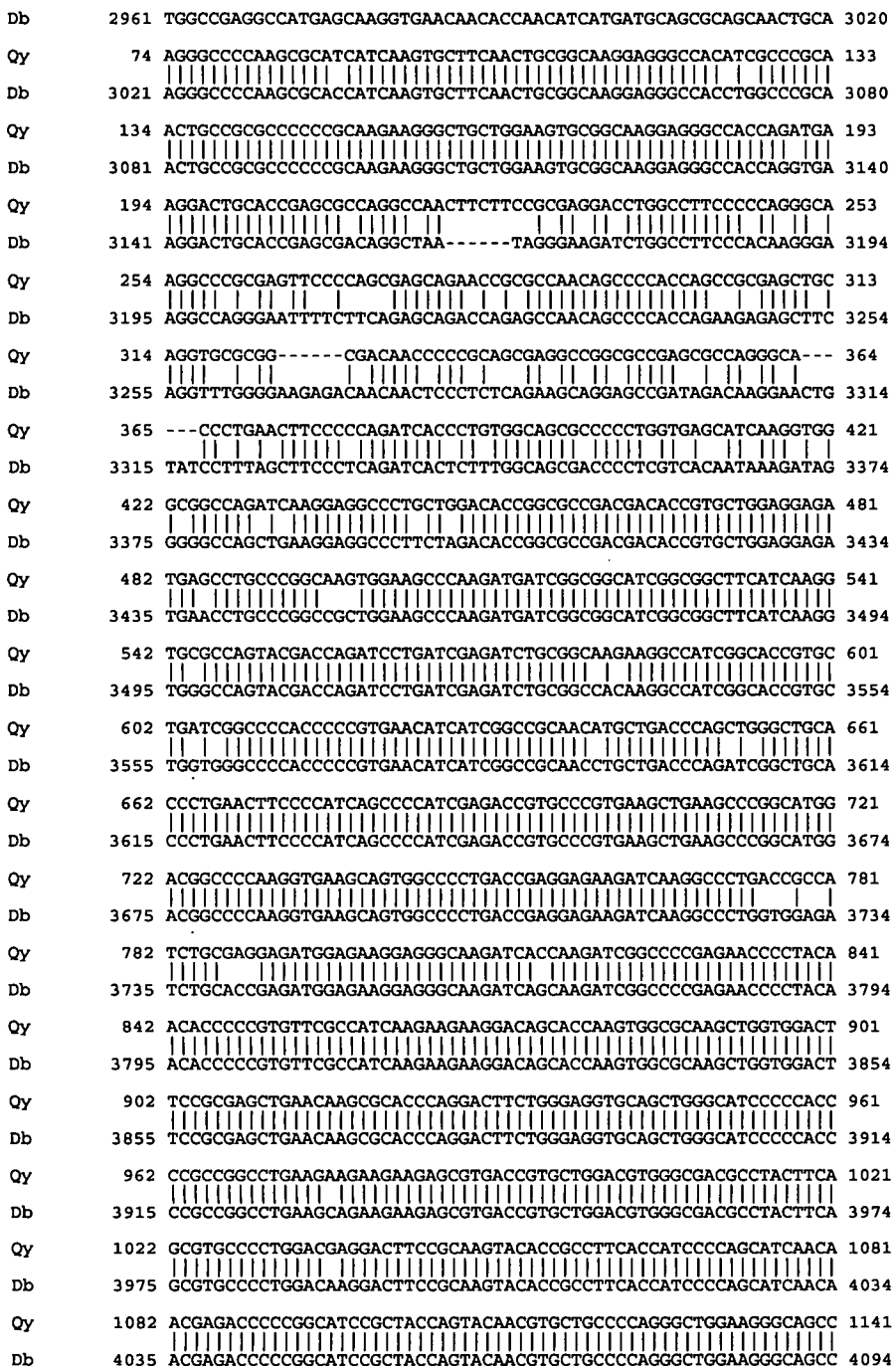
Query Match 82.0%; Score 2019.2; DB 6; Length 2300;
Best Local Similarity 93.2%; Pred. No. 2e-164;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
                |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
                |||||||||||||||  |||||||||||||||||||||  |||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
                |||||||||| |||||||||||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
                |||||||||| |||||||||||||  ||      |||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
                |||||||||||| ||||| | | ||||||||| |||||||||  ||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
                ||||  |||||  |||||||||||||||||  |||||||||||||||||    | ||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
                || | |||||||||||||||||| | ||||||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
                || |||||||||| | ||||||||||||||||||||||||||||||||||||||| ||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
                |||||||||||||||| || |||||||||||||||| ||||||||||||||||||||||
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
```

```
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877

Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997

Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297

Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1357

Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1198

Qy       1358 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1417

Db       1199 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1254

Qy       1418 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477

Db       1255 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1314

Qy       1478 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1537

Db       1315 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1374

Qy       1538 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1597

Db       1375 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1434

Qy       1598 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1657

Db       1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1494

Qy       1658 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1717

Db       1495 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1554

Qy       1718 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1777

Db       1555 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1614

Qy       1778 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1837

Db       1615 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674

Qy       1838 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
```

```
Db       1675 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1734

Qy       1898 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1957

Db       1735 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1794

Qy       1958 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2017

Db       1795 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1854

Qy       2018 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2077

Db       1855 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1914

Qy       2078 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137

Db       1915 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974

Qy       2138 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2197

Db       1975 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2034

Qy       2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257

Db       2035 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2094

Qy       2258 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317

Db       2095 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2154

Qy       2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2377

Db       2155 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2214

Qy       2378 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2437

Db       2215 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2274

Qy       2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463

Db       2275 TTCCCGGGGCTAGCACCGGTGAATTC 2300
```

```
RESULT 14
AX427936
LOCUS       AX427936               9788 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 174 from Patent WO0232943.
ACCESSION   AX427936
VERSION     AX427936.1  GI:21538023
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 174 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9788
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match            82.0%;  Score 2019.2;  DB 6;  Length 9788;
  Best Local Similarity  91.6%;  Pred. No. 1.2e-164;
  Matches 2190;  Conservative    0;  Mismatches  178;  Indels   24;  Gaps    4;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
```


Db 2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020
Qy 74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db 3021 AGGGCCCCAAGCGCACCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3080
Qy 134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db 3081 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGGTGA 3140
Qy 194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db 3141 AGGACTGCACCGAGCGACAGGCTAA------TAGGGAAGATCTGGCCTTCCCACAAGGGA 3194
Qy 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db 3195 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3254
Qy 314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db 3255 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3314
Qy 365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db 3315 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3374
Qy 422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db 3375 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3434
Qy 482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db 3435 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3494
Qy 542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db 3495 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3554
Qy 602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db 3555 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3614
Qy 662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db 3615 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3674
Qy 722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db 3675 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3734
Qy 782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db 3735 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3794
Qy 842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db 3795 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3854
Qy 902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db 3855 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3914
Qy 962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db 3915 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3974
Qy 1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db 3975 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4034
Qy 1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db 4035 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4094
Qy 1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCCGCAACCCCG 1201
Db 4095 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4154
Qy 1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db 4155 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4214
Qy 1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
Db 4215 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4274
Qy 1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1375
Db 4275 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4334
Qy 1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1435
Db 4335 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4394
Qy 1436 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1495
Db 4395 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4454
Qy 1496 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1555
Db 4455 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4514
Qy 1556 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1615
Db 4515 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4574
Qy 1616 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1675
Db 4575 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4634
Qy 1676 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1735
Db 4635 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4694
Qy 1736 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1795
Db 4695 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4754
Qy 1796 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1855
Db 4755 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4814
Qy 1856 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1915
Db 4815 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4874
Qy 1916 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1975
Db 4875 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4934
Qy 1976 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2035
Db 4935 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4994
Qy 2036 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2095
Db 4995 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5054
Qy 2096 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2155
Db 5055 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5114
Qy 2156 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2215
Db 5115 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5174

```
Qy       2216 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2275
Db       5175 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5234

Qy       2276 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2335
Db       5235 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5294

Qy       2336 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db       5295 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5346


RESULT 15
AX427931
LOCUS       AX427931               9169 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 169 from Patent WO0232943.
ACCESSION   AX427931
VERSION     AX427931.1  GI:21538018
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 169 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9169
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             81.4%;  Score 2005.2;  DB 6;  Length 9169;
  Best Local Similarity   91.4%;  Pred. No. 2e-163;
  Matches 2187;  Conservative    0;  Mismatches  183;  Indels   22;  Gaps    5;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
Db       2967 TGGCCGAGGCCATGAGCCAGGTGCAGCACACCAACATCATGATGCAGCGCGGCAACTTCA 3026

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db       3027 AGGGCCAGAAGCGCAT---CAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3083

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
Db       3084 ACTGCCGCGCCCTGCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 3143

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
Db       3144 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3202

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db       3203 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3262

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
Db       3263 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3322

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
Db       3323 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3382

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
Db       3383 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3442

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
Db       3443 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3502

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db       3503 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3562

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db       3563 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3622

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db       3623 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3682

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db       3683 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3742

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db       3743 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3802

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db       3803 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3862

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db       3863 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3922

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db       3923 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3982

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db       3983 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4042

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db       4043 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4102

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db       4103 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4162

Qy       1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db       4163 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4222

Qy       1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
Db       4223 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4282

Qy       1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 1375
Db       4283 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4342

Qy       1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1435
Db       4343 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4402

Qy       1436 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1495
Db       4403 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4462

Qy       1496 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1555
Db       4463 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4522

Qy       1556 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1615
```

```
Db       4523 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4582

Qy       1616 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1675

Db       4583 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4642

Qy       1676 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1735

Db       4643 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4702

Qy       1736 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1795

Db       4703 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4762

Qy       1796 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1855

Db       4763 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4822

Qy       1856 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1915

Db       4823 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4882

Qy       1916 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1975

Db       4883 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4942

Qy       1976 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2035

Db       4943 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 5002

Qy       2036 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2095

Db       5003 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5062

Qy       2096 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2155

Db       5063 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5122

Qy       2156 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2215

Db       5123 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5182

Qy       2216 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2275

Db       5183 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5242

Qy       2276 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2335

Db       5243 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5302

Qy       2336 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387

Db       5303 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5354
```

Search completed: December 30, 2005, 19:16:26
Job time : 11965.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 1296.67 Seconds
(without alignments)
12659.489 Million cell updates/sec

Title: US-09-610-313B-31
Perfect score: 2463
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2463

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2463	100.0	2463	6	ABL39960	Abl39960 Synthetic
2	2463	100.0	2463	12	ADM73765	Adm73765 HIV-1 pol
3	2455.4	99.7	2457	8	ACA03547	Aca03547 Synthetic
4	2455.4	99.7	2457	10	ADC13265	Adc13265 DNA of HI
5	2442.2	99.2	2469	6	ABL39959	Abl39959 Synthetic
6	2442.2	99.2	2469	12	ADM73764	Adm73764 HIV-1 pol
7	2436.2	98.9	2457	6	ABL39961	Abl39961 Synthetic
8	2436.2	98.9	2457	12	ADM73766	Adm73766 HIV-1 pol
9	2430.2	98.7	2457	8	ACA03548	Aca03548 Synthetic
10	2430.2	98.7	2457	10	ADC13266	Adc13266 DNA of HI
11	2422.6	98.4	2445	8	ACA03546	Aca03546 Synthetic
12	2422.6	98.4	2445	10	ADC13264	Adc13264 DNA of HI
13	2415.6	98.1	3930	10	ADC13230	Adc13230 DNA of HI
14	2414	98.0	3930	10	ADC13231	Adc13231 DNA of HI
15	2414	98.0	3930	10	ADC13232	Adc13232 DNA of HI
16	2414	98.0	5184	8	ACA03591	Aca03591 Synthetic
17	2414	98.0	5184	10	ADC13279	Adc13279 DNA of HI
18	2383.6	96.8	3531	10	ADC13234	Adc13234 DNA of HI
19	2382	96.7	3537	10	ADC13236	Adc13236 DNA of HI
20	2381	96.7	5145	8	ACA03521	Aca03521 Synthetic
21	2381	96.7	5145	10	ADC13233	Adc13233 DNA of HI
22	2371	96.3	3538	10	ADC13235	Adc13235 DNA of HI
23	2322.6	94.3	3607	8	ACA03551	Aca03551 Synthetic
24	2322.6	94.3	3607	10	ADC13269	Adc13269 DNA of HI
25	2322.6	94.3	3624	8	ACA03550	Aca03550 Synthetic
26	2322.6	94.3	3624	10	ADC13268	Adc13268 DNA of HI
27	2304.4	93.6	3597	8	ACA03549	Aca03549 Synthetic
28	2304.4	93.6	3597	10	ADC13267	Adc13267 DNA of HI
29	2142	87.0	2466	8	ACA03542	Aca03542 Synthetic
30	2142	87.0	2466	8	ACC78506	Acc78506 HIV p2Pol
31	2122.8	86.2	2472	8	ACA03543	Aca03543 Synthetic
32	2122.8	86.2	2472	8	ACC78507	Acc78507 HIV p2Pol
33	2115.2	85.9	2460	8	ACA03541	Aca03541 Synthetic
34	2115.2	85.9	2460	8	ACC78505	Acc78505 HIV p2Pol
35	2114.2	85.8	3564	8	ACC78488	Acc78488 HIV GagPo
36	2114.2	85.8	3564	8	ACC78489	Acc78489 HIV GagPo
37	2113.6	85.8	4716	8	ACA03522	Aca03522 Synthetic
38	2113.6	85.8	4716	10	ADC13238	Adc13238 DNA of HI
39	2109.8	85.7	3999	8	ACC78484	Acc78484 HIV GagCo
40	2108.2	85.6	3999	8	ACC78485	Acc78485 HIV GagCo
41	2108.2	85.6	3999	8	ACC78486	Acc78486 HIV GagCo
42	2108.2	85.6	5283	8	ACA03584	Aca03584 Synthetic
43	2108.2	85.6	5283	8	ACC78529	Acc78529 HIV TatRe
44	2108	85.6	4713	8	ACA03592	Aca03592 Synthetic
45	2108	85.6	4713	10	ADC13280	Adc13280 DNA of HI

ALIGNMENTS

RESULT 1
ABL39960
ID ABL39960 standard; DNA; 2463 BP.
XX
AC ABL39960;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975YM SEQ ID NO:31.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 9; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef

```
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 2463;  DB 6;  Length 2463;
  Best Local Similarity  100.0%;  Pred. No. 4.4e-297;
  Matches 2463;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
```

```
Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2461 TTC 2463
             |||
Db       2461 TTC 2463
```

RESULT 2
ADM73765
ID ADM73765 standard; DNA; 2463 BP.
XX
AC ADM73765;
XX
DT 03-JUN-2004 (first entry)
XX
DE HIV-1 polynucleotide #8.
XX
KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW HIV type C protein; immunostimulant.
XX
OS Human immunodeficiency virus 1.
XX
PN US2003223961-A1.
XX
PD 04-DEC-2003.
XX
PF 05-JUL-2001; 2001US-00899575.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (MEGE/) MEGEDE J Z.
PA (BARN/) BARNETT S W.
PA (ENGE/) ENGELBRECHT S.
PA (RENS/) RENSBURG E J V.
XX
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;
XX
DR WPI; 2004-060515/06.
XX
PT New expression cassette comprising a polynucleotide sequence encoding an
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT immunization, generating of packaging cell lines or in producing HIV Type
PT C proteins.
XX
PS Claim 1; SEQ ID NO 31; 160pp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV Type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
XX
SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 100.0%; Score 2463; DB 12; Length 2463;
Best Local Similarity 100.0%; Pred. No. 4.4e-297;
Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2461 TTC 2463
              |||
Db       2461 TTC 2463
```

RESULT 3
ACA03547
ID ACA03547 standard; DNA; 2457 BP.
XX
AC ACA03547;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #30.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.

XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 35; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 99.7%; Score 2455.4; DB 8; Length 2457;
Best Local Similarity 100.0%; Pred. No. 3.9e-296;
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 4
ADC13265
ID   ADC13265 standard; DNA; 2457 BP.
XX
AC   ADC13265;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YM_C SEQ ID NO 44.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 41; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match             99.7%;  Score 2455.4;  DB 10;  Length 2457;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-296;
  Matches 2456;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy        61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy       121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy       181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy       241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy       301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy       361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy       421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
```

```
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
```

```
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 5
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.

XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 8; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 99.2%; Score 2442.2; DB 6; Length 2469;
Best Local Similarity 99.6%; Pred. No. 1.7e-294;
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              ||||||||||||||||||      || |  ||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
```

```
Qy       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1915 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2455 GGTGAATTC 2463
             |||||||||
Db       2461 GGTGAATTC 2469
```

RESULT 6
ADM73764
ID ADM73764 standard; DNA; 2469 BP.
XX
AC ADM73764;
XX
DT 03-JUN-2004 (first entry)
XX
DE HIV-1 polynucleotide #7.
XX
KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW HIV type C protein; immunostimulant.
XX
OS Human immunodeficiency virus 1.
XX
PN US2003223961-A1.
XX
PD 04-DEC-2003.
XX
PF 05-JUL-2001; 2001US-00899575.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (MEGE/) MEGEDE J Z.
PA (BARN/) BARNETT S W.
PA (ENGE/) ENGELBRECHT S.
PA (RENS/) RENSBURG E J V.
XX
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;
XX
DR WPI; 2004-060515/06.
XX
PT New expression cassette comprising a polynucleotide sequence encoding an
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT immunization, generating of packaging cell lines or in producing HIV Type
PT C proteins.
XX
PS Claim 1; SEQ ID NO 30; 160pp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV Type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
XX
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 99.2%; Score 2442.2; DB 12; Length 2469;
Best Local Similarity 99.6%; Pred. No. 1.7e-294;
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
              |||||||||||||||||      || |  ||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260

Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320

Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380

Qy       1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440

Qy       1435 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Qy       1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560

Qy       1555 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620

Qy       1615 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680

Qy       1675 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740

Qy       1735 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800

Qy       1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860

Qy       1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920

Qy       1915 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980

Qy       1975 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040

Qy       2035 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100

Qy       2095 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160

Qy       2155 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

Qy       2215 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2275 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2335 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2395 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2455 GGTGAATTC 2463
              |||||||||
Db       2461 GGTGAATTC 2469


RESULT 7
ABL39961
ID   ABL39961 standard; DNA; 2457 BP.
```

XX
AC ABL39961;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975YMWM SEQ ID NO:32.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 10; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 98.9%; Score 2436.2; DB 6; Length 2457;
Best Local Similarity 99.6%; Pred. No. 9.3e-294;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy     1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Db     1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy    61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db    61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy   121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db   121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy   181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db   181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy   241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db   241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy   301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db   301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy   361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db   361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy   421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db   421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy   481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db   481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy   541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db   541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy   601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db   601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy   661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db   661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy   721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db   721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy   781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db   781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy   841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db   841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy   901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db   901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy   961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db   961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy  1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db  1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy  1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db  1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy  1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db  1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy  1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db  1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy  1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db  1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy  1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db  1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374

Qy  1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
```

```
Db       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914

Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154

Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454

Qy       2461 TTC 2463
             |||

Db       2455 TTC 2457
```

```
RESULT 8
ADM73766
ID   ADM73766 standard; DNA; 2457 BP.
XX
AC   ADM73766;
XX
DT   03-JUN-2004  (first entry)
XX
DE   HIV-1 polynucleotide #9.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
PD   04-DEC-2003.
XX
PF   05-JUL-2001; 2001US-00899575.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (MEGE/) MEGEDE J Z.
PA   (BARN/) BARNETT S W.
PA   (ENGE/) ENGELBRECHT S.
PA   (RENS/) RENSBURG E J V.
XX
PI   Megede JZ,  Barnett SW,  Engelbrecht S,  Rensburg EJV;
XX
DR   WPI; 2004-060515/06.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding an
PT   HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT   immunization, generating of packaging cell lines or in producing HIV Type
PT   C proteins.
XX
PS   Claim 1; SEQ ID NO 32; 160pp; English.
XX
CC   The invention relates to an expression cassette comprising a
CC   polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC   also relates to a recombinant expression system for use in a host cell
CC   comprising an expression cassette, where the polynucleotide sequence
CC   further comprises control elements capable of driving expression in the
CC   selected host cell, a cell comprising an expression cassette where the
CC   polynucleotide sequence further comprises control elements compatible
CC   with the expression in the cell and a composition for generating an
CC   immunological response, comprising an expression cassette. The expression
CC   cassette and the methods of the invention are useful in eliciting an
CC   immune response, in DNA immunisation, in generation of packaging cell
CC   lines and in producing HIV Type C proteins. This sequence represents an
CC   HIV-1 polynucleotide of the invention.
XX
SQ   Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

  Query Match             98.9%;  Score 2436.2;  DB 12;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 9.3e-294;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
```

```
Qy       181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db       181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy       241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db       241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy       301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db       301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy       361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db       361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy       421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db       421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db       481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db       541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db       961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy      1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db      1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374

Qy      1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db      1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434

Qy      1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db      1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494

Qy      1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db      1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554

Qy      1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db      1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614

Qy      1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db      1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674

Qy      1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db      1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734

Qy      1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db      1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794

Qy      1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db      1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854

Qy      1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db      1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914

Qy      1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db      1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974

Qy      1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db      1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034

Qy      2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Db      2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094

Qy      2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Db      2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154

Qy      2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Db      2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214

Qy      2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Db      2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274

Qy      2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db      2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334

Qy      2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
```

```
Db       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454

Qy       2461 TTC 2463
              |||
Db       2455 TTC 2457


RESULT 9
ACA03548
ID   ACA03548 standard; DNA; 2457 BP.
XX
AC   ACA03548;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #31.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 36; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

  Query Match             98.7%;  Score 2430.2;  DB 8;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 5.2e-293;
  Matches 2448;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 10
ADC13266
ID ADC13266 standard; DNA; 2457 BP.
XX
AC ADC13266;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 42; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 98.7%; Score 2430.2; DB 10; Length 2457;
Best Local Similarity 99.6%; Pred. No. 5.2e-293;

```
  Matches 2448;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC  66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC  60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              |||||||||||      || |  ||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
```

```
Qy       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 11
ACA03546
ID ACA03546 standard; DNA; 2445 BP.
XX
AC ACA03546;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #29.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 34; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match   98.4%; Score 2422.6; DB 8; Length 2445;
  Best Local Similarity 99.6%; Pred. No. 4.5e-292;
  Matches 2441; Conservative   0; Mismatches   4; Indels   6; Gaps   1;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
```

```
                                                                                                                    
Db        961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020
                                                                                                                    
Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
                                                                                                                    
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080
                                                                                                                    
Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
                                                                                                                    
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
                                                                                                                    
Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
                                                                                                                    
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
                                                                                                                    
Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
                                                                                                                    
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260
                                                                                                                    
Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
                                                                                                                    
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320
                                                                                                                    
Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
                                                                                                                    
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACC 1374
                                                                                                                    
Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
                                                                                                                    
Db       1375 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1434
                                                                                                                    
Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
                                                                                                                    
Db       1435 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1494
                                                                                                                    
Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
                                                                                                                    
Db       1495 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1554
                                                                                                                    
Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
                                                                                                                    
Db       1555 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1614
                                                                                                                    
Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
                                                                                                                    
Db       1615 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1674
                                                                                                                    
Qy       1687 ATGTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
                                                                                                                    
Db       1675 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1734
                                                                                                                    
Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
                                                                                                                    
Db       1735 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1794
                                                                                                                    
Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
                                                                                                                    
Db       1795 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1854
                                                                                                                    
Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
                                                                                                                    
Db       1855 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1914
                                                                                                                    
Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
                                                                                                                    
Db       1915 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1974
                                                                                                                    
Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
                                                                                                                    
Db       1975 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2034
                                                                                                                    
Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
                                                                                                                    
Db       2035 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2094
                                                                                                                    
Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
                                                                                                                    
Db       2095 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2154
                                                                                                                    
Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
                                                                                                                    
Db       2155 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2214
                                                                                                                    
Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
                                                                                                                    
Db       2215 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2274
                                                                                                                    
Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
                                                                                                                    
Db       2275 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2334
                                                                                                                    
Qy       2347 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2406
                                                                                                                    
Db       2335 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2394
                                                                                                                    
Qy       2407 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
                                                                                                                    
Db       2395 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 12
ADC13264
ID ADC13264 standard; DNA; 2445 BP.
XX
AC ADC13264;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2Pol-opt-YMWM_C SEQ ID NO 43.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 40; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,

CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 98.4%; Score 2422.6; DB 10; Length 2445;
Best Local Similarity 99.6%; Pred. No. 4.5e-292;
Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

```
Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy        727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy        787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy        847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy        907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy        967 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy       1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy       1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy       1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy       1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy       1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy       1327 CACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACC 1386
              |||||||||||||||||||||||  ||      ||||||||||||||||||||||||||
Db       1321 CACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACC 1374

Qy       1387 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1434

Qy       1447 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1494

Qy       1507 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1554

Qy       1567 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1614

Qy       1627 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1674

Qy       1687 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1734

Qy       1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1794

Qy       1807 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1854

Qy       1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1914

Qy       1927 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1974

Qy       1987 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2034
```

```
Qy       2047 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2094

Qy       2107 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2154

Qy       2167 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2155 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2214

Qy       2227 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2215 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2274

Qy       2287 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2275 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2334

Qy       2347 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2335 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2394

Qy       2407 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2395 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

RESULT 13
ADC13230
ID ADC13230 standard; DNA; 3930 BP.
XX
AC ADC13230;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagComplPolmut_C SEQ ID NO 9.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 6; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

```
  Query Match             98.1%;  Score 2415.6;  DB 10;  Length 3930;
  Best Local Similarity   99.6%;  Pred. No. 3.1e-291;
  Matches 2434;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
```

```
Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy      1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db      2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy      1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db      2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy      1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
Db      2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy      1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
Db      2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy      1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
Db      2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980

Qy      1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
Db      2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy      1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
Db      3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy      1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
Db      3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy      1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
Db      3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220

Qy      1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
Db      3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy      1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
Db      3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy      1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
Db      3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy      1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
Db      3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy      1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
Db      3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy      2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
Db      3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy      2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
Db      3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy      2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
Db      3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700

Qy      2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
Db      3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy      2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
Db      3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy      2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
Db      3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy      2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db      3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 14
ADC13231
ID ADC13231 standard; DNA; 3930 BP.
XX
AC ADC13231;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagComplPolmutAtt_C SEQ ID NO 10.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 7; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a

```
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;

  Query Match             98.0%;  Score 2414;  DB 10;  Length 3930;
  Best Local Similarity   99.5%;  Pred. No. 4.9e-291;
  Matches 2433;  Conservative    0;  Mismatches    5;  Indels    6;  Gaps    1;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
               | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
              |||||||||||||||    ||       ||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860

Qy       1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920

Qy       1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980

Qy       1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040

Qy       1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100

Qy       1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160

Qy       1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220

Qy       1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280

Qy       1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340

Qy       1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400

Qy       1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460

Qy       1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520

Qy       2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580

Qy       2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640

Qy       2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700

Qy       2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760

Qy       2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820

Qy       2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880

Qy       2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
                |||||||||||||||||||||||||||||||||||||||||||
Db       3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

RESULT 15
ADC13232
ID ADC13232 standard; DNA; 3930 BP.
XX
AC ADC13232;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagComplPolmutIna_C SEQ ID NO 11.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 8; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 98.0%; Score 2414; DB 10; Length 3930;
Best Local Similarity 99.5%; Pred. No. 4.9e-291;
Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
```

```
Db      2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326
Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db      2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386
Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446
Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506
Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566
Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626
Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
Qy      1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db      2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746
Qy      1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db      2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806
Qy      1334 AGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGC 1393
Db      2807 AGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGC 2860
Qy      1394 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 1453
Db      2861 CCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCA 2920
Qy      1454 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
Db      2921 AGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980
Qy      1514 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 1573
Db      2981 TGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGC 3040
Qy      1574 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 1633
Db      3041 TGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCA 3100
Qy      1634 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 1693
Db      3101 AGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACC 3160
Qy      1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 1753
Db      3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCA 3220
Qy      1754 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 1813
Db      3221 ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCT 3280
Qy      1814 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 1873
Db      3281 GGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGA 3340
Qy      1874 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 1933
Db      3341 CCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGG 3400
Qy      1934 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993
Db      3401 TGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3460
Qy      1994 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 2053
Db      3461 ACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC 3520
Qy      2054 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 2113
Db      3521 GGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCC 3580
Qy      2114 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 2173
Db      3581 AGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCC 3640
Qy      2174 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 2233
Db      3641 TGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCG 3700
Qy      2234 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 2293
Db      3701 AGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCG 3760
Qy      2294 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353
Db      3761 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGG 3820
Qy      2354 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 2413
Db      3821 ACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCG 3880
Qy      2414 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db      3881 GCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

Search completed: December 30, 2005, 08:57:01
Job time : 1304.67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:41 ; Search time 8697.67 Seconds
(without alignments)
13249.138 Million cell updates/sec

Title: US-09-610-313B-31
Perfect score: 2463
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2463

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	136.4	5.5	330	10	CL293849	CL293849 02S0349-0
	2	98.4	4.0	2886	10	CL967755	CL967755 OsIFCC015
	3	90.2	3.7	1781	8	CX099623	CX099623 RECm0260
	4	86.8	3.5	2598	4	AY103647	AY103647 Zea mays
	5	86.6	3.5	1398	10	CL961989	CL961989 OsIFCC006
	6	86	3.5	743	10	CZ247380	CZ247380 AIAA-aaf3
	7	85.6	3.5	951	3	BM321451	BM321451 rockefell
	8	85.4	3.5	2031	10	CL974989	CL974989 OsIFCC042
	9	84	3.4	1725	10	CL978463	CL978463 OsIFCC031
c	10	80.8	3.3	1060	10	CW922203	CW922203 EDCAR29TR
	11	80.8	3.3	3069	10	CL973991	CL973991 OsIFCC025
c	12	80.6	3.3	869	7	CK159167	CK159167 FGAS04056
	13	80.6	3.3	892	10	CZ216254	CZ216254 AIAA-aaf2
	14	80.6	3.3	1941	10	CL971508	CL971508 OsIFCC021
	15	80.2	3.3	1132	3	BM320864	BM320864 rockefell
	16	79.6	3.2	1165	3	BM320900	BM320900 rockefell
	17	79.6	3.2	11691	10	CL962901	CL962901 OsIFCC008
	18	79	3.2	867	3	BM321430	BM321430 rockefell
	19	78.6	3.2	2682	10	CL969033	CL969033 OsIFCC017
	20	78.4	3.2	873	7	CO443531	CO443531 MZCCL1006
	21	78.2	3.2	2853	10	CL974397	CL974397 OsIFCC025
	22	78	3.2	1485	10	CL970981	CL970981 OsIFCC020
	23	78	3.2	2313	10	CL982362	CL982362 OsIFSC047
	24	78	3.2	2454	10	CL975440	CL975440 OsIFCC027
	25	77.4	3.1	1509	10	CL959255	CL959255 OsIFCC002
	26	77.4	3.1	2559	10	CL982027	CL982027 OsIFSC046
	27	76.6	3.1	1550	3	BM321022	BM321022 rockefell
	28	76.4	3.1	2028	10	CL979437	CL979437 OsIFCC033
	29	76.2	3.1	1401	10	CL962721	CL962721 OsIFCC038
	30	76	3.1	1680	10	CL982770	CL982770 OsIFSC049
	31	76	3.1	2793	10	CL948321	CL948321 OsIFSB004
	32	75	3.0	862	3	BM321023	BM321023 rockefell
	33	75	3.0	1734	10	CL948553	CL948553 OsIFSB005
	34	74.6	3.0	1410	10	CL947314	CL947314 OsIFSB004
	35	74.6	3.0	1989	10	CL980973	CL980973 OsIFCC035
	36	74.4	3.0	914	9	BZ568300	BZ568300 pacs2-164
c	37	74	3.0	1071	10	CW927105	CW927105 EDCBJ95TF
	38	74	3.0	1290	10	CL972679	CL972679 OsIFCC023
	39	74	3.0	1386	11	DQ045165	DQ045165 Homo sapi
	40	74	3.0	2072	4	CR603312	CR603312 full-leng
	41	73.2	3.0	853	3	BM321393	BM321393 rockefell
	42	73	3.0	2151	10	CL972100	CL972100 OsIFCC041
c	43	72.8	3.0	566	3	BM587428	BM587428 170006873
	44	72.6	2.9	788	6	CB643171	CB643171 OSJNEb03L
	45	72.4	2.9	753	9	CC675888	CC675888 OGWCO51TH

ALIGNMENTS

RESULT 1
CL293849/c
LOCUS CL293849 330 bp DNA linear GSS 12-FEB-2004
DEFINITION 02S0349-08A1-C03 UniformMu MuTAIL Library Zea mays genomic clone 02S0349-08A1-C03, genomic survey sequence.
ACCESSION CL293849
VERSION CL293849.1 GI:42541978
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 330)
AUTHORS Latshaw,S., Tan,B-.C., Settles,A.M. and McCarty,D.R.
TITLE Sequence tagged transposon insertions from the UniformMu maize population
JOURNAL Unpublished (2003)
COMMENT Contact: Donald R. McCarty
Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drm@ufl.edu
Sequence flanking probable Mu insertion site in UniformMu line: 02S0349-08, Primer set: A
Class: transposon insertion site.
FEATURES Location/Qualifiers
source 1. .330
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="02S0349-08A1-C03"
/clone_lib="UniformMu MuTAIL Library"
/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo assymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sepharose 400 spin columns and cloned into the TOPO PCR4 vector."
ORIGIN

```
  Query Match             5.5%;  Score 136.4;  DB 10;  Length 330;
  Best Local Similarity   66.4%;  Pred. No. 2.7e-14;
  Matches  211;  Conservative    0;  Mismatches  106;  Indels    1;  Gaps    1;

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              || || ||| |  | || || ||  | || || || || || || || ||  | || ||
Db        329 GGGGGGCAGCTGGAAGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAA 270

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||   || | || |  ||||| || || ||||| || || || || || || |||||
Db        269 ATGACTTTGACAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAA 210

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||  | ||||| || ||| | |  || || ||||| ||  | || || || || || |||
Db        209 GTAAGACAGTATGATCAGGTACCCATAGAAATCTGTGGGCATAAAGCTATAGGTACGGTA 150

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              |  | || || || || || ||||| || ||  | ||  || |||| ||| | || |||
Db        149 TTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGG-TGC 91

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||| | || || ||||| || || || || || || || || ||  | ||||| || |||
Db         90 ACCTTAAATTTTCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATG 31

Qy        721 GACGGCCCCAAGGTGAAG 738
              || ||||| || || | |
Db         30 GATGGCCCAAAAGTAAGG 13


RESULT 2
CL967755
LOCUS       CL967755               2886 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC015718 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL967755
VERSION     CL967755.1  GI:52390149
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2886)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .2886
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             4.0%;  Score 98.4;  DB 10;  Length 2886;
  Best Local Similarity   43.2%;  Pred. No. 2.1e-07;
  Matches  709;  Conservative    0;  Mismatches  906;  Indels   27;  Gaps    4;

Qy        189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
              | | | |||   | |||   |  ||||| | || ||  | | | |   | |  || |

Db        219 GGTCATGGAGGCCGCCGCCATCATGGCCATCGTCCTCGCCAACGGCGGCGGCAGGCCGCC 278

Qy        249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
              | | |    |   ||  || ||       || || || || | ||||| | |
Db        279 CGACTGGCAGGACTTCGTCGGCATCGTCGTCCTCCTCGTCATCAACTCCACCATCAGTTT 338

Qy        309 GCTGCAGGTGCGC---GGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCAC 365
              |  ||| |  |   | ||  ||| ||  |   | |   |  |||| | |  |  ||
Db        339 CATCGAGGAGAACAATGCCGGGAACGCCGCCGCGGCGCTCATGGCCGGCCTCGCGCCCAA 398

Qy        366 CCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGG 425
              ||  | | |  |  ||||||  |   | |  ||||  ||  |||
Db        399 GACCAAGGTGCTCAGGGACGGGAAATGGCAGGAGCAGGACGCGTCCATCCTCGTTCCCGG 458

Qy        426 CCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAG 485
              | | ||||    |  ||| | |  |  |  || |  || || |  |  |  | | |
Db        459 CGACATCATCAGCATCAAGCTCGGCGACATCATCCCCGCCGACGCGCGGCTGCTCGAGGG 518

Qy        486 CCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCG 545
              |  ||  ||||   | |  |  | || ||||      |  ||||  ||
Db        519 CGACCCGCTCAAGGTCGACCAGGCGGCGCTCACCGGCGAGTCGATGCCCGTCAACAAGCA 578

Qy        546 CCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGAT 605
              |       |  || |  || | |  || |||| ||| |      |  |  |  |
Db        579 CGCCGGGCAGGGCGTCTTCTCCGGGTCCACCGTCAAGCAGGGCGAGATCGAGGCCGTCGT 638

Qy        606 CGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCT 665
              |  | |||||  |||| ||| | ||  | ||||  |  |  || | ||| | |||||
Db        639 CATCGCCACCGGCGTGCACACCTTCTTCGGCAAGGCGGCGCACCTGGTGGACAGCACCAA 698

Qy        666 GAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGG 725
              ||| ||  |  ||  ||  ||| |  |  |||  |                |  ||| |
Db        699 CAACATCGGCCACTTCCAGCTCGTGCTCACGGCCATCGGCAACTTCTGCATCATCTCCAT 758

Qy        726 CCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTG 785
              |  |  | | |  |  | | ||    | |  |  |    |  |||  | |
Db        759 CGGCGTCGGCATGATAATCGAGATCATCGTTATGTACCCGATCCAGCACCGCGCGTACCG 818

Qy        786 CGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACAC 845
              ||| |  || || ||   |  |   |  ||   ||||  |  | | ||     |||
Db        819 CGACGGCATCGACAACCTGCTCGTCCTGCTCATCGGCGGCATCCCCATCGCCATGCCCAC 878

Qy        846 CCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCG 905
              |   ||| || || || |  |  |  |  |  |  |  |  ||| || | | ||
Db        879 CGTGCTGTCCGTCACCATGGCCATCGGCTCCCACCGGCTGTCGCAGCAGGGCGCCATCAC 938

Qy        906 CGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGC 965
              | ||| ||  |   || | ||||  |   |  || |  ||  |  |  | |||  |
Db        939 CAAGCGGATGACCGCCATCGAGGAGATGGCCGGCATGGACGTGCTCTGCAGCGACAAGAC 998

Qy        966 CGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGT 1025
              |||   |  || |  || |  | |||||   |   |   ||     | | |
Db        999 CGGGACGCTGACGCTCAACAAGCTCACCGTCGACAAGACGCTGATCGAGGTGTACGGGCG 1058

Qy       1026 GCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGA 1085
              |  |||||| ||||||              |||   |       || || || | ||| |
Db       1059 GGGGCTGGACAAGGACTCGGTGCTCCTGTACGCGGCGAGGGCGTCCCGCGTCGAGAACCA 1118

Qy       1086 GACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAG 1145
              |  | ||   |  | |         | |||| || |   |  | ||| | || || |
Db       1119 GGACGCCATTGACACATGCATCGTGGGCATGCTCGCCGACCCCAAGGAGGCCCGCGCCGG 1178

Qy       1146 CATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGAT 1205
              ||||    ||  |  | |   | ||   | ||| ||      ||||   ||   |
Db       1179 CATCAAGGAGGTCCACTTCCTCCCGTTC---AACCCCGTCGAGAAGCGCACGGCCATCAC 1235

Qy       1206 CGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGC 1265
              |  ||| ||          | | | | | ||          ||||  |
Db       1236 CTACATCGACGGCAATGGCGAATGGCACAGGATCAGCAAGGGCGCGCCGGAGCAGATCAT 1295

Qy       1266 CAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAA 1325
              | ||  |  || || ||||| ||   |  ||   |||      |||| || |
Db       1296 CGAGCTGTGCAAGATGAGCAAGGACGCCGAGAAGAAGGTGCACACGCTGATCGACCAGTA 1355
```

```
Qy       1326 GCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGAC 1385
Db       1356 CGCGGACCGCGGCCTCCGGTCGCTGGGCGTGTCGTACCAGAAGGTGCCGGAGAAG----- 1410

Qy       1386 CGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCT 1445
Db       1411 ----------AGCAAGGAGAGCGAGGGCGAGCCATGGCAGTTCGTCGGCCTCCTGCCGCT 1460

Qy       1446 GGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTG 1505
Db       1461 GTTCGACCCGCCGAGGCACGACAGCGCGGAGACCATCCGCCGCGCGCTGCACCTCGGCGT 1520

Qy       1506 CAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGA 1565
Db       1521 GAACGTGAAGATGATCACCGGCGACCAGCTCGCCATCG------GGAAGGAGACGGCGCG 1574

Qy       1566 GCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGA 1625
Db       1575 GCGGCTCGGGATGGGCACCAACATGTACCCGTCGACGACGCTGCTGGGCGACAAGAGCAG 1634

Qy       1626 CCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCA 1685
Db       1635 CGAGATGAGCGGCCTCCCAATCGACGAGCTGATCGAGAAGGCCGACGGCTTCGCCGGCGT 1694

Qy       1686 GATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGC 1745
Db       1695 GTTCCCGGAGCACAAGTACGAGATCGTGAAGCGCCTCCAGGACCGGAAGCACATCTGCGG 1754

Qy       1746 CCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCAT 1805
Db       1755 CATGACCGGCGACGGCGTGAACGACGCGCCGGCGCTGAAGAAGGCCGACATCGGCATCGC 1814

Qy       1806 CGTGATCTGGGGCAAGACCCCC 1827
Db       1815 CGTCGACGACGCCACGGACGCC 1836
```

```
RESULT 3
CX099623
LOCUS       CX099623               1781 bp    mRNA    linear   EST 03-JUN-2005
DEFINITION  RECm0260 A normalized whole-life-cycle cDNA library of rice Oryza
            sativa (indica cultivar-group) cDNA clone EI057O14, THREEB09,
            EI057O04, EI062D05, BI027I14, EI070G16, EI08 5', mRNA sequence.
ACCESSION   CX099623
VERSION     CX099623.1  GI:66912775
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1781)
  AUTHORS   Zhang,J., Peng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D.,
            Han,B., Zhang,Q. and Wang,S.
  TITLE     Features of the expressed sequences revealed by a large-scale
            analysis of ESTs from a normalized cDNA library of the elite indica
            rice cultivar Minghui 63
  JOURNAL   Plant J. 42 (5), 772-780 (2005)
   PUBMED   15918889
COMMENT     Contact: Wang S
            National Key Laboratory of Crop Genetic Improvement
            Hauzhong Agricultural University
            Wuhan 430070, China
            Tel: 86-27-87282044
            Fax: 86-27-87287092
            Email: shipingwang@hotmail.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .1781
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /strain="indica"
                     /cultivar="Minghui 63"
                     /db_xref="taxon:39946"
                     /clone="EI057O14, THREEB09, EI057O04, EI062D05, BI027I14,
                     EI070G16, EI08"
                     /tissue_type="whole plant"
                     /dev_stage="whole-life-cycle"
                     /lab_host="E. coli DH10B"
                     /clone_lib="A normalized whole-life-cycle cDNA library of
                     rice"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
                     library is constructed based on the strategy of satuation
                     hybridization with genomic DNA using rice cultivar Minghui
                     63. This library consists of cDNA from 15 directionally
                     cloned cDNA libraries constructed with different tissues
                     from 9 developmental stages."
ORIGIN
```

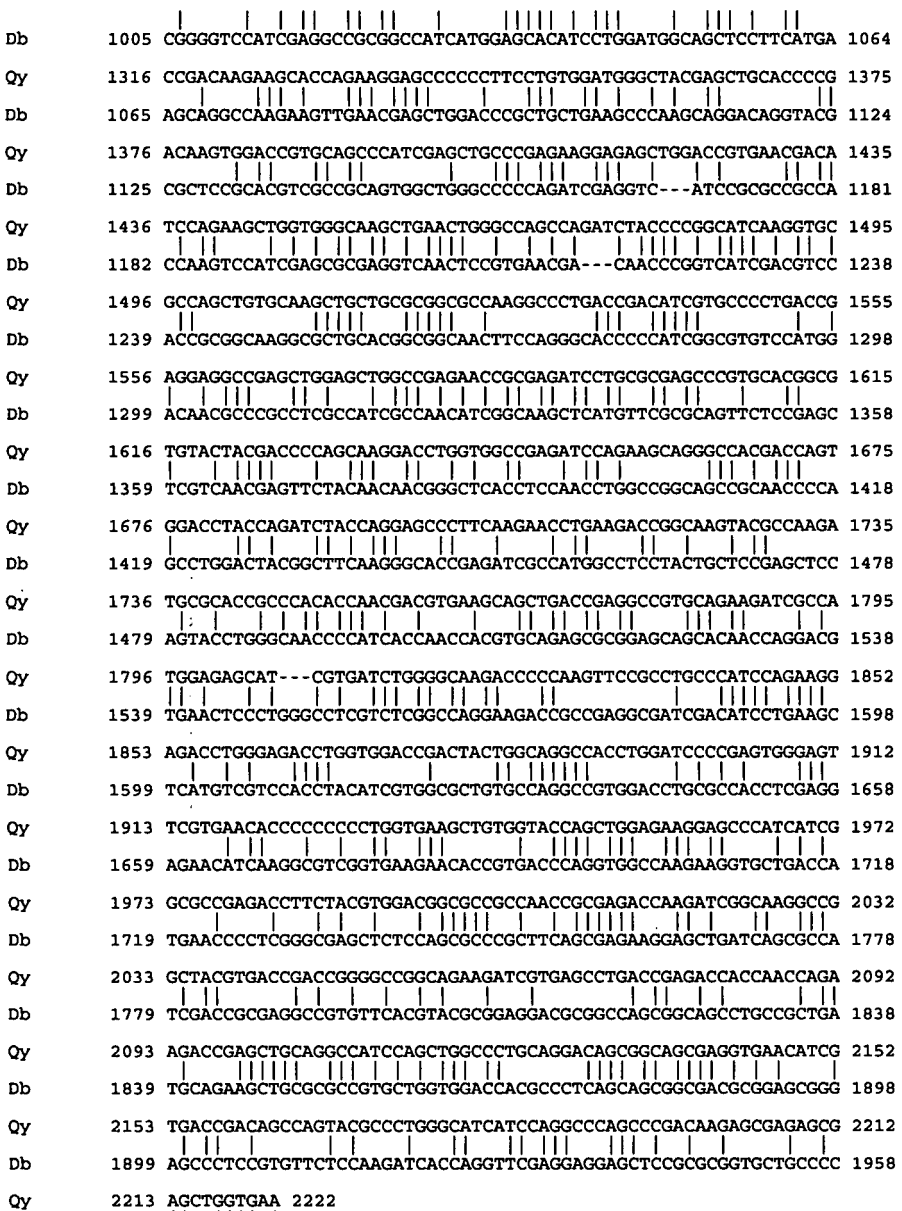
```
  Query Match             3.7%;  Score 90.2;  DB 8;  Length 1781;
  Best Local Similarity   43.8%;  Pred. No. 6e-06;
  Matches  447;  Conservative    0;  Mismatches  568;  Indels    6;  Gaps    1;

Qy       231 GGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAA 290
Db       194 GGACGCGGTGCTCGACGCGTGCCTCGCCCAGGACCCCGACAGCAAGGTGGCTTGCGAGAC 253

Qy       291 CAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGC 350
Db       254 GTGCACCAAGACCAACATGGTGATGGTGTTCGGCGAGATCACCACCAAGGCCACCGTCGA 313

Qy       351 CGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAG 410
Db       314 CTACGAGAAGATCGTCCGCGACACCTGCCGCGGCATCGGCTTCGTGTCCGACGACGTCGG 373

Qy       411 CATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGT 470
Db       374 CCTCGACGCCGACCGCTGCAAGGTGCTCGTCAACATCGAGCAGCAGTCGCCCGACATCGC 433

Qy       471 GCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGG 530
Db       434 GCAGGGGGTGCACGGCCACTTCACCAAGCGCCCCGAGGAGATCGGCGCCGGCGACCAGGG 493

Qy       531 CTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCAT 590
Db       494 CCACATGTTCGGCTACGCCACCGACGAGACCCCCGAGCTGATGCCCCTCAGCCACGTCCT 553

Qy       591 CGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCA 650
Db       554 CGCCACCAAGCTCGGCGCCCGCCTCACCGAGGTCCGCAAGAACGGCACCTGCGCGTGGCT 613

Qy       651 GCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAA 710
Db       614 CAGGCCCGACGGCAAGACCCAGGTCACCGTCGAGTACCTCAACGACGCCGGCGCCATGGT 673

Qy       711 GCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGC 770
Db       674 CCCCGTCCGCGTCCACACCGTCCTCATCTTCACCCAGCACGACGAGACCGTCACCAACGA 733

Qy       771 CCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGA 830
Db       734 CGAGATCGCCGCC------GACCTCAAGGAGCACGTCATCAAGCCGGTCATCCCCGACAA 787

Qy       831 GAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAA 890
Db       788 GTACCTCGACGAGAAGACCATCTTCCACCTCAACCCCTCGGGCCGCTTCGTCATCGGCGG 847

Qy       891 GCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGG 950
Db       848 GCCCCACGGCGACGCCGGCCTCACCGGCCGCAAGATCATCATCGACACCTACGGCGGATG 907

Qy       951 CATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGA 1010
Db       908 GGGCGCGCACGGCGGCGGCGCCTTCTCCGGCAAGGACCCGACCAAGGTCGACCGGAGCGG 967
```

```
Qy       1011 CGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCC 1070
Db        968 CGCCTACATCGCCAGGCAGGCCGCCAAGAGCATCGTCGCCAGCGGCCTCGCCCGCCGCTG 1027

Qy       1071 CAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTG 1130
Db       1028 CATCGTGCAGGTGTCGTACGCCATCGGCGTCCCCGAGCCGCTCTCCGTGTTCGTCGACTC 1087

Qy       1131 GAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGC 1190
Db       1088 CTACGGCACCGGCAAGATCCCCGACAAGGAGATCCTCAAGATCGTCAAGGAGAACTTCGA 1147

Qy       1191 CCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGAT 1250
Db       1148 TTTCAGGCCCGGGATGATGACCATCAACCTCGACCTCAAGAGGGGCGGCAACCGGTTCAT 1207

Qy       1251 C 1251
Db       1208 C 1208
```

RESULT 4
AY103647
LOCUS AY103647 2598 bp mRNA linear HTC 18-FEB-2005
DEFINITION Zea mays PCO142084 mRNA sequence.
ACCESSION AY103647
VERSION AY103647.1 GI:21206725
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2598)
AUTHORS Gardiner,J., Schroeder,S., Polacco,M.L., Sanchez-Villeda,H., Fang,Z., Morgante,M., Landewe,T., Fengler,K., Useche,F., Hanafey,M., Tingey,S., Chou,H., Wing,R., Soderlund,C. and Coe,E.H. Jr.
TITLE Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization
JOURNAL Plant Physiol. 134 (4), 1317-1326 (2004)
PUBMED 15020742
REFERENCE 2 (bases 1 to 2598)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 3 (bases 1 to 2598)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES Location/Qualifiers
source 1. .2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeGDB:638378"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
ORIGIN

Query Match 3.5%; Score 86.8; DB 4; Length 2598;
Best Local Similarity 42.1%; Pred. No. 2.5e-05;
Matches 813; Conservative 0; Mismatches 1102; Indels 15; Gaps 5;

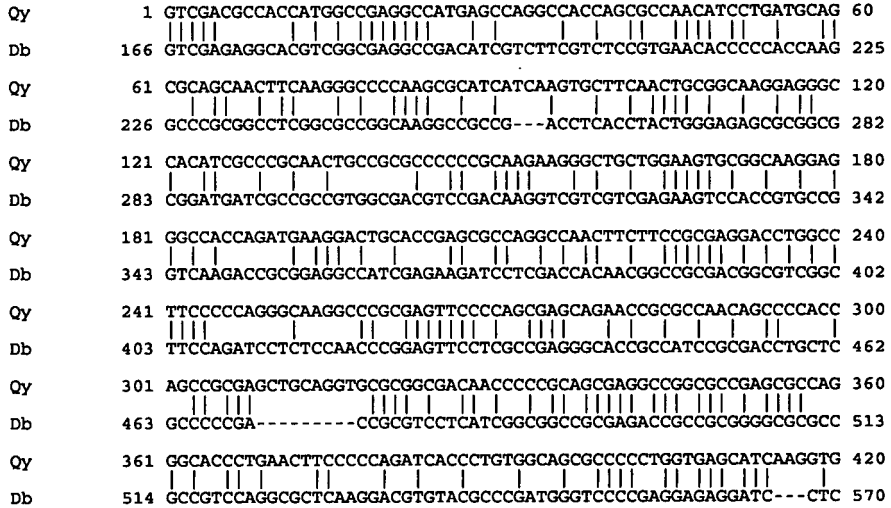
```
Qy        296 CCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGC 355
Db         51 CCACCACCTCCCACCTCCTGCCACCGCGCGGCCCCAACCAACCACACCGCAGCGCAGCAA 110

Qy        356 GCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCA 415
Db        111 TGGCGGGCAACGGCGCCATCGTGGAGAGCGACCCGCTGAACTGGGGCGCGGCGGCGGCGG 170

Qy        416 AGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGG 475
Db        171 AGCTGGCCGGGAGCCACCTGGACGAGGTGAAGCGCATGGTGGCGCAGGCCCGGCAGCCCG 230

Qy        476 AGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCA 535
Db        231 TGGTCAAGATCGAGGGTTCCACCCTCCGCGTCGGCCAGGTGGCCGCCGTCGCCTCCGCCA 290

Qy        536 TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCA 595
Db        291 AGGACGCGTCCGGCGTCGCCGTCGAGCTCGACGAGGAGGCCCGCCCCCGCGTCAAGGCCA 350

Qy        596 CCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGG 655
Db        351 GCAGCGAGTGGATCCTCGACTGCATCGCCCACGGCGGCGACATCTACGGCGTCACCACCG 410

Qy        656 GCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCG 715
Db        411 GCTTCGGCGGCACCTCCCACCGCCGCACCAAGGACGGGCCCGCGCTCCAGGTCGAGCTGC 470

Qy        716 GCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGA 775
Db        471 TCAGGCATCTCAACGCCGGAATCTTCGGCACCGGCAGCGACGGGCACACGCTGCCGTCGG 530

Qy        776 CCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACC 835
Db        531 AGGTCACCCGCGCGGCGATGCTGGTGCGCATCAACACCCTCCTCCAGGGCTACTCCGGCA 590

Qy        836 CCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGG 895
Db        591 TCCGCTTCGAGATCCTCGAGGCCATCACGAAGCTGCTCAACACCGGTGTCAGCCCCTGCC 650

Qy        896 TGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCC 955
Db        651 TGCCGCTCCGGGGCACCATCACCGCGTCGGGCGACCTGGTCCCGCTCTCCTACATCGCCG 710

Qy        956 CCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCT 1015
Db        711 GCCTCATCACGGGCCGCCCCAACGCGCAGGCCGTCACCG---TCGACGGGAGGAAGGTGG 767

Qy       1016 ACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCA 1075
Db        768 ACGCCGCCGAGGCGTTCAAGATCGCCGGCATCGAGGGCGGCTTCTTCAAGCTCAACCCCA 827

Qy       1076 TCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGG 1135
Db        828 AGGAGGGCCTCGCCATCGTCAACGGCACGTCCGTGGGCTCCGCGCTCGCGGCCACCGTGA 887

Qy       1136 GCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCA 1195
Db        888 TGTACGACGCCAACGTCCTGGCCGTCCTGTCGGAGGTCCT---GTCCGCCGTCTTCTGCG 944

Qy       1196 ACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db        945 AGGTCATGAACGGCAAGCCCGAGTACACGGACCACCTGACCCACAAGCTGAAGCACCACC 1004

Qy       1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
```

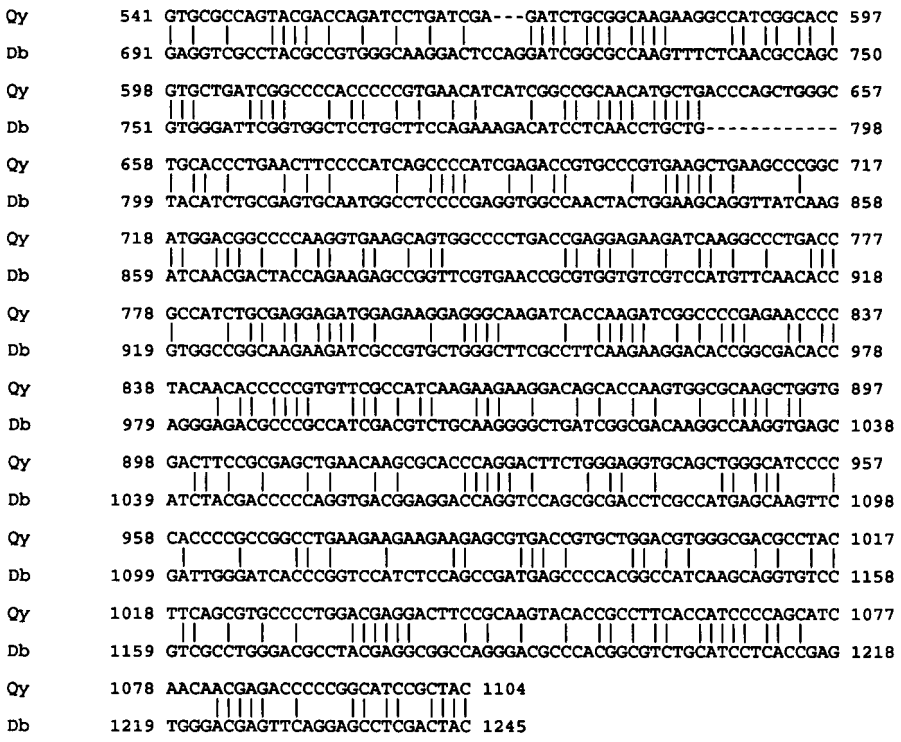


RESULT 5
CL961989
LOCUS CL961989 1398 bp DNA linear GSS 21-SEP-2004
DEFINITION OsIFCC006991 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.

ACCESSION CL961989
VERSION CL961989.1 GI:52378720
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 1398)
AUTHORS Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES Location/Qualifiers
source 1. .1398
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN

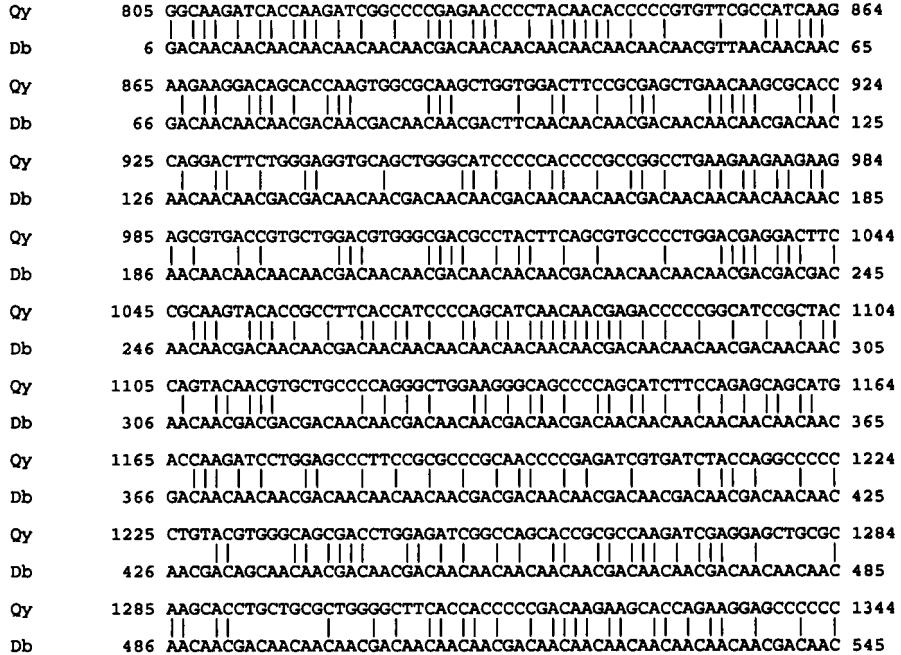
Query Match 3.5%; Score 86.6; DB 10; Length 1398;
Best Local Similarity 45.9%; Pred. No. 2.6e-05;
Matches 508; Conservative 0; Mismatches 569; Indels 30; Gaps 5;





```
RESULT 6
CZ247380
LOCUS       CZ247380                743 bp    DNA     linear   GSS 11-FEB-2005
DEFINITION  AIAA-aaf35a12.b1 Ancylostoma caninum whole genome shotgun library
            (AIAAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION   CZ247380
VERSION     CZ247380.1  GI:59632821
KEYWORDS    GSS.
SOURCE      Ancylostoma caninum (dog hookworm)
  ORGANISM  Ancylostoma caninum
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
            Ancylostomatoidea; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
REFERENCE   1  (bases 1 to 743)
  AUTHORS   Mitreva,M., McCarter,J.P., Pape,D., Ritter,E., Tsagareishvili,R.,
            Ronko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Messina,D.,
            Waterston,R.H., Clifton,S.W. and Wilson,R.
  TITLE     Genome Survey sequences from the parasitic nematode Ancylostoma
            caninum
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Mitreva M
            Washington University in St. Louis
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: nematode@watson.wustl.edu
            Genomic DNA provided by John Hawdon (mtmjmh@gwumc.edu) DNA
            sequenced by Washington University Genome Sequencing Center
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .743
                     /organism="Ancylostoma caninum"
                     /mol_type="genomic DNA"
                     /strain="Baltimore"
                     /db_xref="taxon:29170"
                     /dev_stage="Adult"
                     /lab_host="GS10"
                     /clone_lib="Ancylostoma caninum whole genome shotgun
                     library (AIAAGSS 001)"
                     /note="Vector: pOTW13; Site_1: BstX1; Site_2: BstX1;
                     Ancylostoma caninum genomic DNA was randomly sheared,
                     end-repaired and size fractioned to enrich for 2-4 kb
                     fragments. Genomic DNA was provided by John Hawdon
                     (mtmjmh@gwumc.edu) at George Washington University.
                     Sequencing by Washington University Genome Sequencing
                     Center, St. Louis, MO."
ORIGIN
```

Query Match 3.5%; Score 86; DB 10; Length 743;
Best Local Similarity 45.9%; Pred. No. 3.2e-05;
Matches 293; Conservative 0; Mismatches 345; Indels 0; Gaps 0;



```
RESULT 7
BM321451
LOCUS       BM321451                951 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
            3.3.1.1), mRNA sequence.
ACCESSION   BM321451
VERSION     BM321451.1  GI:18055857
```

```
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 951)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 951   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .951
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.5%;  Score 85.6;  DB 3;  Length 951;
  Best Local Similarity   46.0%;  Pred. No. 3.9e-05;
  Matches  323;  Conservative    0;  Mismatches  376;  Indels    3;  Gaps    1;

Qy       1728 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1787
Db        219 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 278

Qy       1788 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1847
Db        279 CATCGCGCAGGCCGGCGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 338

Qy       1848 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1907
Db        339 GGAGTGCACCTGGAAGGCCCTGTGCTTCGGCCCCTACCAGGGCCCTCAGATCATCGTCGA 398

Qy       1908 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1967
Db        399 CGACGGCGGTGACGCGACTCTGATGATCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 458

Qy       1968 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2027
Db        459 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 518

Qy       2028 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2087
Db        519 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 578

Qy       2088 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2147
Db        579 CGAGGAGACGACGACTGGCGTNATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 638

Qy       2148 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2207
Db        639 GTTCCCGGCCGTCAACGTCAACGAC---TCTNTCACCAAGAGCAAGTTCGACAACATCTA 695

Qy       2208 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2267
Db        696 CGGCTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGG 755

Qy       2268 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2327
Db        756 CAAGGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCG 815

Qy       2328 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2387
Db        816 CGGCCAGGGCTGCCGCGTCATCGTGACGGAGATCGACCCCATCTGCGCGCTGCAGGCGTC 875

Qy       2388 CATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGA 2429
Db        876 GATGGCCGGCTTCGAGGTCAACACGCTCGAGGCGGGGCTCGA 917


RESULT 8
CL974989
LOCUS       CL974989               2031 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC042895 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL974989
VERSION     CL974989.1  GI:52404497
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2031)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .2031
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.5%;  Score 85.4;  DB 10;  Length 2031;
  Best Local Similarity   42.6%;  Pred. No. 4.4e-05;
  Matches  787;  Conservative    0;  Mismatches 1026;  Indels   36;  Gaps    5;

Qy        576 CGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCG 635
Db        153 CGGCACGACGTACTCGTGCGTCGGAGTCTACCGGAACGGCCATGTCGAGATCATCGCCAA 212

Qy        636 CAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGT 695
Db        213 CGACCAGGGGAACCGGATCACGCCCTCGTGGGTCGCCTTCACCGACGGCGGCGAGCGGCT 272

Qy        696 GCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGA 755
Db        273 CATCGGTGAGGCCGCCAAGAACCAGGCGGCGGCCAACCCGGAGCGGACCATCTACGACGC 332

Qy        756 GGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCAC 815
Db        333 CAAGCGGCTCATCGGCCGGCAGTTCTCCGACGCCGAGGTGCAGCGCGACATGAAGCTGCT 392

Qy        816 CAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAG 875
Db        393 GCCGTTCGCCGTCGTCGACCGGAACGGCAAGCCGCACGTGCGCGTCGAGGTGAAGGACGG 452

Qy        876 CACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTG 935
```

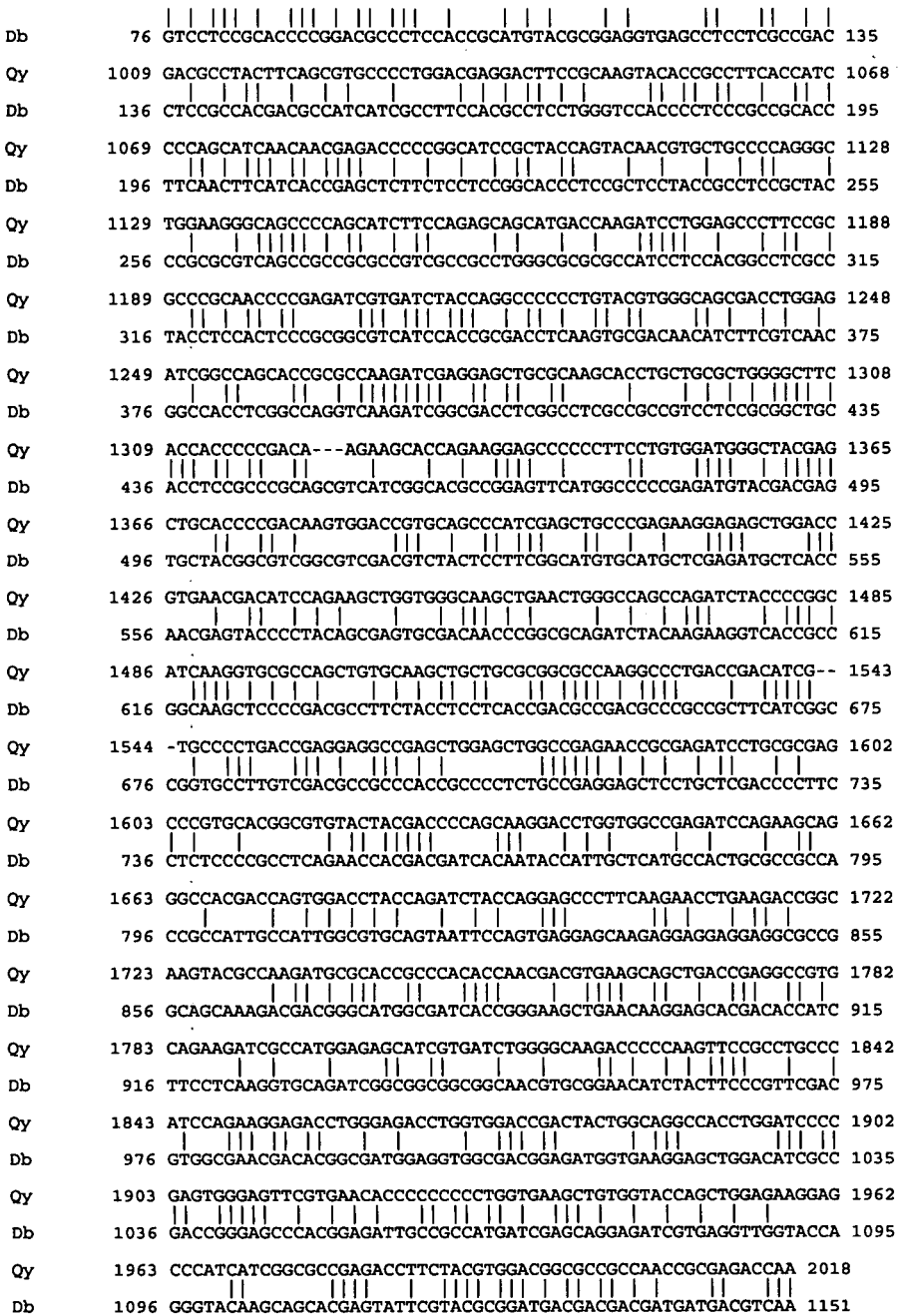
```
Db        453 CGACGTGCG-----------CGTGTTCAGCCCGGAGGAGGTGAGCGCCATGGTGCTCAC 500
Qy        936 GGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGT 995
Db        501 GCGGATGAAGGAGACGGCCGAGGCCTACCTCGGCGAGAAGGTCACCCGCGCCGTCGTCAC 560
Qy        996 GCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACAC 1055
Db        561 CGTCCCGGCCTACTTCAACGACGCGCAGCGGCAGGCCACCAAGGACGCCGGCGTCATCGC 620
Qy       1056 CGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGT 1115
Db        621 CGGGCTCACCGTCGACCGCATCATCAACGAGCCCACCGCCGCCGCCATCGCCTACGGCAT 680
Qy       1116 GCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175
Db        681 CGACAAGAAGGGCGCCGAGAAGAACGTCCTCGTCTTCGACCTCGGCGGCGGCACGTTCGA 740
Qy       1176 GGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTG-- 1233
Db        741 CGTCAGCATCCTCGCCATCGACAACGGCGTGTTCGAGGTCCTTGCCACCAACGGCGACAC 800
Qy       1234 ----GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCA 1289
Db        801 CCACCTCGGCGGCGAGGACTTCGACCAACGCCTCATGGACCACTTCGTCAAGGTCATCCG 860
Qy       1290 CCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCT 1349
Db        861 CCGGAAGCACGGCCGCGACATCACCGGCGACGCGCGCGCGCTGGGCAAGCTCCGCCGCGA 920
Qy       1350 GTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCT---GCC 1406
Db        921 GTGCGAGCGCGCCAAGCGCGCGCTCAGCAACCAGCACCAGGTGCGCGTCGAGGTCGAGTC 980
Qy       1407 CGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGC 1466
Db        981 CCTGTTCGACGGCGTCGACTTGTCGGAGCCGCTCTCCCGGGCGCGGTTCGAGGAGCTCAA 1040
Qy       1467 CAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAA 1526
Db       1041 CAGCGACCTCTTCAAGAAGACGATGGTGCCGGTGAGGAAGGCCATGGCGGACGCCCGGCT 1100
Qy       1527 GGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCG 1586
Db       1101 GAGCAAGGGCGACATCGACGAGATCGTCCTCGTCGGCGGCAGCACCAGGATCCCCAAGGT 1160
Qy       1587 CGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGC 1646
Db       1161 GCAGCAGCTGCTCAAGGACTACTTCGGTGGCAAGGAGCCCAACCGCGGCGTCAACCCCGA 1220
Qy       1647 CGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAA 1706
Db       1221 CGAGGCCGTGGCGTACGGCGCCGCCGTGCAGGCCAGCATCATCAGCGGCCACGTCGACGA 1280
Qy       1707 GAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCA 1766
Db       1281 GAACACGGAGAGCATGATCCTCCTCGACGTCGCGCCGCTCACCCTCGGCCTGGAGACGGC 1340
Qy       1767 GCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCC 1826
Db       1341 CGGCGGCGTGATGGCCAAGCTGATCCCGCGCAACACGGTGGTGCCGACGAAGAAGACGCA 1400
Qy       1827 CAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCA 1886
Db       1401 GGTGTTCACCACGTACAAGGACAAGCAGACAACGGTGACCATCCAGGTGTTCGAGGGCGA 1460
Qy       1887 GGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTA 1946
Db       1461 GCGGAGCATGACGAGGGACAACCGGCTGCTCGGCAGGTTCGACCTCGCCGGCATCGCGCC 1520
Qy       1947 CCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAA 2006
Db       1521 GGCGCCGAGGGGGGCGCCGCAGATCGAGGTGACG---------TTCGAGGTGGACGCGAA 1571
```

```
Qy       2007 CCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGT 2066
Db       1572 CGGCATCCTCAGCGTGCTGGCCGCCGACAAGGCCACCGGCAGGTCGGAGAAGATCACCAT 1631
Qy       2067 GAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCA 2126
Db       1632 CTCCGGCGACGACCGCAAGATCAGCCAGGAGGAGATCGACCGGATGGTGCGCGAGGCGGA 1691
Qy       2127 GGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCA 2186
Db       1692 GGAGTTCGCCGAGGAGGACCGCCGGCACAGGGAGCAGGTGGACGCCCGGAACAGCCTGGA 1751
Qy       2187 GGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGAT------CATCGAGCAGCT 2240
Db       1752 GGCGTACGTCTACAACATCAAGAACACGCTCGGCGGCAAGATGGCGGACGCCATGGAGGG 1811
Qy       2241 GATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAA 2300
Db       1812 CGAGGAGAAGGACAAGGTGGAGGAGGCGGTGAGGGAGGCGTACGAGTGGCTGGACGGCAA 1871
Qy       2301 CGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCAT 2360
Db       1872 CCCGGACGCCGGCAAGGAGGAGTACGAGGAGAAGCTGAGGGAGCTGGAGGACGTGTGCAA 1931
Qy       2361 CGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGC 2409
Db       1932 CCCCGTCATGTCGGCGGTCTACCAGAGGTCCGGCGGCGGCGGCGGCGGC 1980
```

RESULT 9
CL978463
LOCUS CL978463 1725 bp DNA linear GSS 21-SEP-2004
DEFINITION OsIFCC031917 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL978463
VERSION CL978463.1 GI:52411427
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 1725)
AUTHORS Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES Location/Qualifiers
source 1. .1725
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Query Match 3.4%; Score 84; DB 10; Length 1725;
Best Local Similarity 43.7%; Pred. No. 7.8e-05;
Matches 470; Conservative 0; Mismatches 600; Indels 6; Gaps 2;

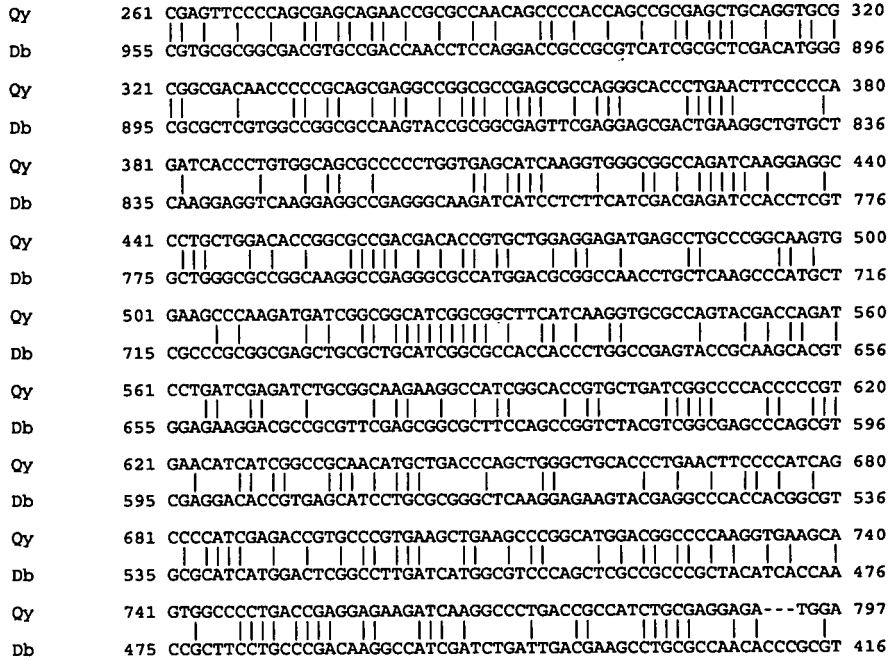
```
Qy        949 GGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGC 1008
```

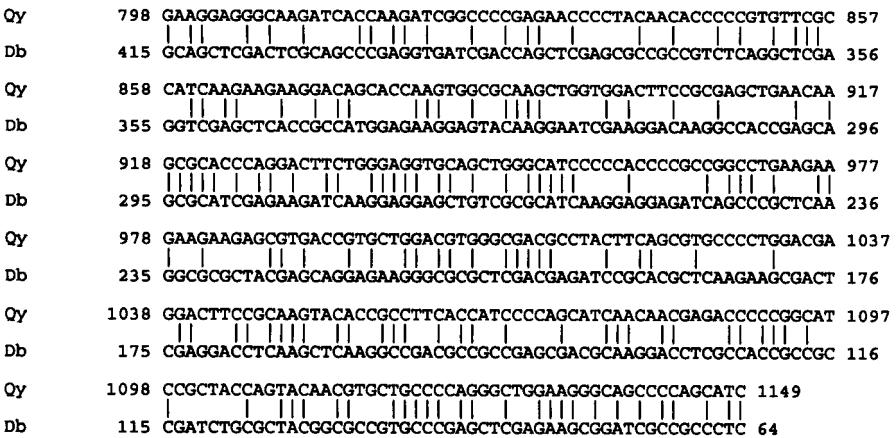


RESULT 10

CW922203/c
LOCUS CW922203 1060 bp DNA linear GSS 24-FEB-2005
DEFINITION EDCAR29TR A. castellanii, 6-8 kb library from total genomic DNA Acanthamoeba castellanii genomic clone EDCAR29, genomic survey sequence.
ACCESSION CW922203
VERSION CW922203.1 GI:60228146
KEYWORDS GSS.
SOURCE Acanthamoeba castellanii
ORGANISM Acanthamoeba castellanii
Eukaryota; Acanthamoebidae; Acanthamoeba.
REFERENCE 1 (bases 1 to 1060)
AUTHORS Anderson,I.J. and Loftus,B.J.
TITLE Gene discovery in the Acanthamoeba castellanii genome
JOURNAL Unpublished (2004)
COMMENT Contact: Iain Anderson
The Institute for Genomic Research (TIGR; www.tigr.org)
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-795-7949
Fax: 301-838-0208
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .1060
/organism="Acanthamoeba castellanii"
/mol_type="genomic DNA"
/strain="Neff"
/db_xref="taxon:5755"
/clone="EDCAR29"
/clone_lib="A. castellanii, 6-8 kb library from total genomic DNA"
/note="Vector: pHOS2"
ORIGIN

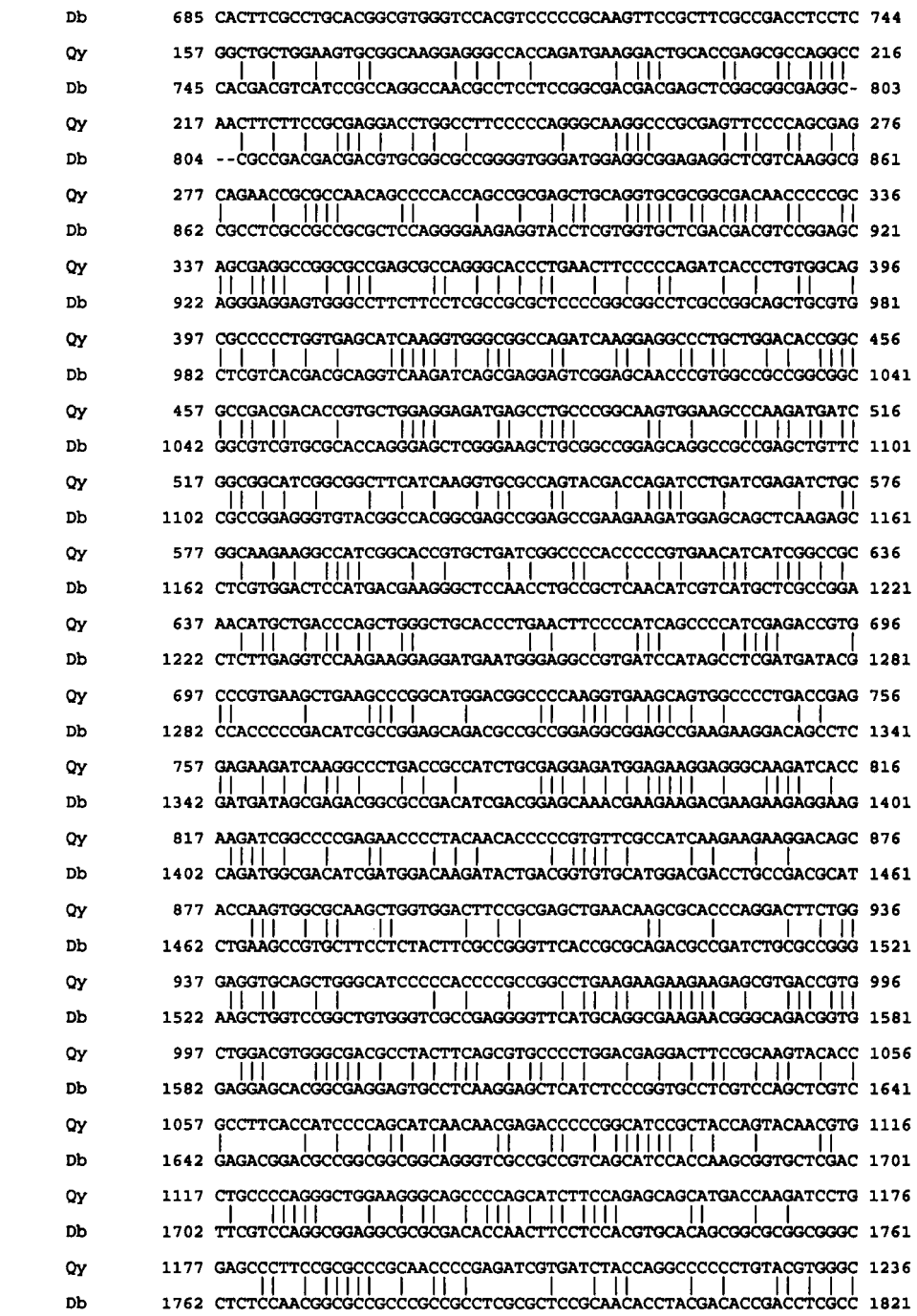
Query Match 3.3%; Score 80.8; DB 10; Length 1060;
Best Local Similarity 43.9%; Pred. No. 0.00028;
Matches 392; Conservative 0; Mismatches 497; Indels 3; Gaps 1;





RESULT 11
CL973991
LOCUS CL973991 3069 bp DNA linear GSS 21-SEP-2004
DEFINITION OsIFCC025416 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL973991
VERSION CL973991.1 GI:52402507
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 3069)
AUTHORS Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES Location/Qualifiers
source 1. .3069
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Query Match 3.3%; Score 80.8; DB 10; Length 3069;
Best Local Similarity 41.6%; Pred. No. 0.0003;
Matches 657; Conservative 0; Mismatches 912; Indels 9; Gaps 2;



```
Qy       1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
Db       1822 GTCATGCTCGAGGCGCCCAAGCTCCACACCCTCCTCTGCGACATCCCGGAGCGCG----- 1876

Qy       1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATG 1356
Db       1877 -CCGCCGCCGACGCCGCCGGCGCCGCGGCGTGGCAGAGAGCGCTCGAGCTCATCAATGGC 1935

Qy       1357 GGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAG 1416
Db       1936 CGCGCCCCCACCTTCTCCGTCCACGGCTCGAGGTTCCTCCGCGTGATGGATCTCAAGGGC 1995

Qy       1417 AGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATC 1476
Db       1996 GTGCGGCTGCCGCACCGGGAGTCGCTGCCGGAGGAGATCGGGTGGCTGATCCACCTCAGG 2055

Qy       1477 TACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACC 1536
Db       2056 TACCTGGGCCTCTCGCACACCACCATGAGGCACCTCCCGAGCTCCGTCAAGAGGCTCCGG 2115

Qy       1537 GACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTG 1596
Db       2116 AACCTGCAGACGCTGGACGTGTCGCGCACGGACGTCGAGGCGCTGCCGTGGCGGCTCTGG 2175

Qy       1597 CGCGAGCCCGTGCACGGC 1614
Db       2176 CGCAACCCCTCGCTCCGC 2193
```

RESULT 12
CK159167/c
LOCUS CK159167 869 bp mRNA linear EST 05-DEC-2003
DEFINITION FGAS040564 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA, mRNA sequence.
ACCESSION CK159167
VERSION CK159167.1 GI:38985053
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 869)
AUTHORS Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D., Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A., Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D., Penniket,C., Roach,J.L. and Sarhan,F.
TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL Unpublished (2003)
COMMENT Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon, Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas_ests@cs.usask.ca
This sequence is the direct result of the Base calling software Phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [128,636].
Plate: TaLt537 row: N column: 23.

FEATURES Location/Qualifiers
source 1. .869
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DH5 alpha"
/clone_lib="Triticum aestivum FGAS: TaLt5"
/note="Organ: Crown; Vector: pGEM-T; SSH (suppression subtractive hybridization) cDNA library from genotype PI178383 cold hardened at 2 C for 21 days and 49 days (equal amount of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 H)(driver). Modified Smart cDNA (Clontech)priming and non-directional cloning"

ORIGIN

Query Match 3.3%; Score 80.6; DB 7; Length 869;
Best Local Similarity 44.9%; Pred. No. 0.0003;
Matches 305; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

```
Qy        580 AAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAAC 639
Db        827 ATGGCGGCGGCCCGCACAACGAGGATCACCAACCAGGACGAGAACAACCACAACCACCAC 768

Qy        640 ATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCC 699
Db        767 CACAACGACGACAACAGCGACACAACGACCACCAACAACACCAACAACAACAACAACAAC 708

Qy        700 GTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAG 759
Db        707 ACCAACAACAACAACAACAACAACAACAACAACGACAACCACCACCACAACCACAACAAC 648

Qy        760 AAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAG 819
Db        647 AACAACAACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 588

Qy        820 ATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACC 879
Db        587 AACAACAACAACAACAACAACAACAACAACAACGACAACAACAACAACAACAACAACAAC 528

Qy        880 AAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAG 939
Db        527 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 468

Qy        940 GTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTG 999
Db        467 AACAACAACAACAACAACAACAACAACGACGACGACGACAACGACGACGACGACGACGAC 408

Qy       1000 GACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCC 1059
Db        407 GACAACAACGACAACAACGACAACAACAACAACGACAACAACAACAACAACAACAACAAC 348

Qy       1060 TTCACCATCCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTG 1119
Db        347 GACCACAACGACAACGACAACAACAACAACAACAACAACAACAACGACAACAACCACAAC 288

Qy       1120 CCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG 1179
Db        287 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 228

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGC 1239
Db        227 AACAACAACAACAACAACGACGACAACGACGTCAACGATAACAACGACAACAACGACAAC 168

Qy       1240 GACCTGGAGATCGGCCAGC 1258
Db        167 AACGGCAACAACAACAAAC 149
```

RESULT 13
CZ216254
LOCUS CZ216254 892 bp DNA linear GSS 10-FEB-2005
DEFINITION AIAA-aaf23g02.b1 Ancylostoma caninum whole genome shotgun library (AIAAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION CZ216254
VERSION CZ216254.1 GI:59229909
KEYWORDS GSS.
SOURCE Ancylostoma caninum (dog hookworm)
ORGANISM Ancylostoma caninum
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida; Ancylostomatoidea; Ancylostomatidae; Ancylostomatinae; Ancylostoma.

```
REFERENCE   1  (bases 1 to 892)
  AUTHORS   Mitreva,M., McCarter,J.P., Pape,D., Ritter,E., Tsagareishvili,R.,
            Ronko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Messina,D.,
            Waterston,R.H., Clifton,S.W. and Wilson,R.
  TITLE     Genome Survey sequences from the parasitic nematode Ancylostoma
            caninum
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Mitreva M
            Washington University in St. Louis
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: nematode@watson.wustl.edu
            Genomic DNA provided by John Hawdon (mtmjmh@gwumc.edu) DNA
            sequenced by Washington University Genome Sequencing Center
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .892
                     /organism="Ancylostoma caninum"
                     /mol_type="genomic DNA"
                     /strain="Baltimore"
                     /db_xref="taxon:29170"
                     /dev_stage="Adult"
                     /lab_host="GS10"
                     /clone_lib="Ancylostoma caninum whole genome shotgun
                     library (AIAAGSS 001)"
                     /note="Vector: pOTW13; Site_1: BstX1; Site_2: BstX1;
                     Ancylostoma caninum genomic DNA was randomly sheared,
                     end-repaired and size fractioned to enrich for 2-4 kb
                     fragments. Genomic DNA was provided by John Hawdon
                     (mtmjmh@gwumc.edu) at George Washington University.
                     Sequencing by Washington University Genome Sequencing
                     Center, St. Louis, MO."
ORIGIN

  Query Match             3.3%;  Score 80.6;  DB 10;  Length 892;
  Best Local Similarity   46.5%;  Pred. No. 0.0003;
  Matches  334;  Conservative    0;  Mismatches  379;  Indels    6;  Gaps    2;

Qy       1650 GATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAA 1709
Db          5 GATTCAGCTCGAGCACCTCCAGCAGCACGTCAACCAGCACGTCAACCACCACCTCGACGT 64

Qy       1710 CCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCT 1769
Db         65 CGACGACACCAAGCACAACCACATCCACAACCACTTCGACGACACCGACGACTACTAGCT 124

Qy       1770 GACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAA 1829
Db        125 CTAGCACGTCAAGCAGCACGTCAACCAGCACGTCAACCAGCACATCGACGTCGACGACAC 184

Qy       1830 GTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGC 1889
Db        185 CAAGCACAACTACGTCCACCACCACTTCGACGACACCGACGACTACTAGCTCTAGCACGT 244

Qy       1890 CACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCA 1949
Db        245 CAAGCAGCACGTCAACCAGCACGTCAACCAGCACCTCGACGTCGACGACACCAAGCCCAA 304

Qy       1950 GCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCG 2009
Db        305 CCACGTCCACAACCACTTCGACGACACCGACGACTACTACCTCGAGTACCTCGACCAGCA 364

Qy       2010 CGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAG 2069
Db        365 CGTCAACCAGCACCTCGACGTCGACGACACCAAGCACAACCACGTCCACAACCACTTCGA 424

Qy       2070 CCTGACCGAGACCACCAA---CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCA 2126
Db        425 CGACACCGACGACTACTAGCTCGAGCACGTCAAGCAGCACGTCAACCAGCACGTCAACCA 484

Qy       2127 GGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCA 2186
Db        485 GCACCTCGACGTCGACGACACCAAGCACAACCACGTCCACAACCACTTCGACGACACCGA 544

Qy       2187 GGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAA 2246
Db        545 CGACTACTAGCTCGAGCACGTCAACCAGCACGTCAACCAGCACGTCGACCAGCACATCGA 604

Qy       2247 GAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCA 2306
Db        605 CGTCGACGACACCAAGCACAACCACGTCCACAACCACTGCGACGACAGCGACGACTACTA 664

Qy       2307 GATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATG 2365
Db        665 GCTCG---AGCACGTCAACCAGCACGTCAACCAGCACCTCGACGTCGACGACACCGAGG 720


RESULT 14
CL971508
LOCUS       CL971508               1941 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC021485 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL971508
VERSION     CL971508.1  GI:52397596
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1941)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .1941
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.3%;  Score 80.6;  DB 10;  Length 1941;
  Best Local Similarity   42.4%;  Pred. No. 0.00032;
  Matches  757;  Conservative    0;  Mismatches 1004;  Indels   24;  Gaps    5;

Qy        267 CCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGA 326
Db        165 CGCCGCCAAGAACCAGGTCGCCATGAACCCAACCAACACCGTCTTCGATGCGAAGCGGTT 224

Qy        327 CAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCAC 386
Db        225 GATTGGTAGGAGGTTCTCCGACCCGTCCGTGCAGAGTGACATGAAGCTATGGCCGTTCAA 284

Qy        387 CCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCT 446
Db        285 GGTGGTGCCTGGCCCCGGCGACAAGCCGATGATCGTCGTCCAGTACAAGGGCGAGGAGAA 344

Qy        447 GGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCC 506
Db        345 GCAGTTCGCCGCGGAGGAGATCTCGTCCATGGTGCTGATCAAGATGAGGGAGATCGCGGA 404
```

```
Qy       507 CAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGAT 566
Db       405 GGCCTACCTCGGCTCGTCCATCAAGAACGCCGTGGTCACCGTCCCGGCCTACTTCAACGA 464

Qy       567 CGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACAT 626
Db       465 CTCGCAGAGGCAGGCCACCAAGGACGCCGGCGTCATCGCCGGGCTCAACGTGATGCGCAT 524

Qy       627 CATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCAT 686
Db       525 CATCAACGAGCCCACCGCCGCCGCCATCGCGTACGGCCTCGACAAGAAGGCGACGAGCAG 584

Qy       687 CGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCC 746
Db       585 CGGCGAGAAGAACGTGCTCATCTTTGACCTCGGCGGCGGCACGTTCGATGTGTCGCTGCT 644

Qy       747 CCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGG 806
Db       645 CACCATCGAGGAGGGCATCTTCGAGGTCAAGGCCACCGCCGGCGACACCCACCTCGGCGG 704

Qy       807 CAAGATCACCAAGATCGGCCCCGAGAACCCCTAC---AACACCCCCGTGTTCGCCATCAA 863
Db       705 CGAGGACTTCGACAACCGCATGGTGAACCACTTCGTGCAGGAGTTCAAGCGCAAGAACAA 764

Qy       864 GAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCAC 923
Db       765 GAAGGACATCAGCGGCAACCCGCGCGCGCTGCGGCGGCTCAGGACGGCGTGCGAGCGCGC 824

Qy       924 CCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAA 983
Db       825 CAAGCGCACGCTGTCGTCGACGGCGCAGACCACCATTGAGATCGACTCGCTGTACGAGGG 884

Qy       984 GAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTT 1043
Db       885 CATCGACTTCTACACGACGATCACCAGGGCGCGGTTCGAGGAGCTCAACATGGACCTGTT 944

Qy      1044 CCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTA 1103
Db       945 CCGCAAGTGCATGGAGCCCGTGGAGAAGTGCCTCCGCGACGCCAAGATGGACAAGAGCAG 1004

Qy      1104 CCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCAT 1163
Db      1005 CGTGCACGACGTCGTGCTCGTCGGCGGCTCCACCCGCATCCCCAAGGTGCAGCAGCTGCT 1064

Qy      1164 GACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCC 1223
Db      1065 CCAGGACTTCTTCAACGGCAAGGAGCTCTGCA------AGAGCATCAACCCCGACGAGGC 1118

Qy      1224 CCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCG 1283
Db      1119 CGTGGCGTACGGCGCCGCCGTCCAGGCCGCCATCCTGAGCGGCGAGGGCAACGAGAAGGT 1178

Qy      1284 CAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCC 1343
Db      1179 GCAGGACCTGCTCCTGCTCGACGTCACCCCGCTGTCCCTCGGCCTCGAGACCGCCGGCGG 1238

Qy      1344 CTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCA---GCCCATCGA 1400
Db      1239 CGTCATGACCGTGCTCATCCCGAGGAACACCACCATCCCGACCAAGAAGGAGCAGGTCTT 1298

Qy      1401 GCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAA 1460
Db      1299 CTCCACCTACTCCGACAACCAGCCCGGCGTCCTGATCCAGGTGTACGAGGGCGAGCGCGC 1358

Qy      1461 CTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA---------AGCT 1511
Db      1359 CAGGACCAAGGACAACAACCTCCTCGGCAAGTTCGAGCTCTCCGGCATCCCACCGGCGCC 1418

Qy      1512 GCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGA 1571
Db      1419 GCGCGGCGTGCCCCAGATCAACGTCTGCTTCGACATCGACGCCAACGGCATCCTGAACGT 1478

Qy      1572 GCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAG 1631
Db      1479 GTCCGCGGAGGACAAGACGACGGGGCAGAAGAACAAGATCACCATCACCAACGACAAGGG 1538

Qy      1632 CAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTA 1691
Db      1539 GCGGCTGAGCAAGGAGGAGATCGAGAAGATGGTGCAGGAGGCGGAGAAGTACAAGGCCGA 1598

Qy      1692 CCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACAC 1751
Db      1599 GGACGAGGAGCACAAGAAGAAGGTGGACGCCAAGAACGCCCTCGAGAACTACGCCTACAA 1658

Qy      1752 CAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGAT 1811
Db      1659 CATGCGCAACACCATCAAGGACGAGAAGATCGCCTCCAAGCTCGCCGCCGACGACAAGAA 1718

Qy      1812 CTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTG 1871
Db      1719 GAGGATCGAGGACGCCATCGATGGCGCCATCAGCTGGCTCGACACCAACCAGCTCGCCGA 1778

Qy      1872 GACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCT 1931
Db      1779 GGCCGACGAGTTCGAGGACA---AGATGAAGGAGCTGGAGGGCATCTGCAACCCGATCAT 1835

Qy      1932 GGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGT 1991
Db      1836 CGCCAAGATGTACCAGGGCCCCGGCGCCGACATGGCCGGAGGCATGGACGAGGACGCGCC 1895

Qy      1992 GGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTA 2036
Db      1896 GGCGGGCGGCAGCGGCGCAGGCCCCAAGATCGAGGAGGTCGACTA 1940
```

```
RESULT 15
BM320864
LOCUS       BM320864               1132 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM320864
VERSION     BM320864.1  GI:18055270
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1132)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1132   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1132
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.3%;  Score 80.2;  DB 3;  Length 1132;
  Best Local Similarity   45.3%;  Pred. No. 0.00037;
  Matches  375;  Conservative    0;  Mismatches  444;  Indels    9;  Gaps    2;
```

```
Qy        139 CGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGAC 198
              |||  |    |||||| || || | |   | ||     || |  |||  | | |||  |
Db         79 CGCCGCGAGGGCAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAG 138

Qy        199 TGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCC 258
              | || |   | | ||  |  |||| ||       | ||    |  |  |  || ||  | |
Db        139 TACAACAGCCCCAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTC 192

Qy        259 CGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG 318
               || || || ||  ||   ||| |  || | ||  |  |  |  ||          |
Db        193 TGCCAGATCGCCTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCAC 252

Qy        319 CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCC 378
              |  |  |  | ||  | ||  ||||        | | | | ||    ||    |
Db        253 GAGCTCACCCGCTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACT 312

Qy        379 CAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAG 438
              |   |||  |  ||   ||| ||   |||| | | |  | ||  | |||
Db        313 GGCCTGCTGCTGGCCCGCCGTGTGCTGAAGAAGCTCAACCTCGACTCCAAGTACGAGGGT 372

Qy        439 GCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAG 498
              | |  |  || || |||||  ||| ||| |||   ||||       |  |||| ||
Db        373 GTCAAGAAGGTCAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGT 432

Qy        499 TGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAG 558
              |  ||| ||  | |   || |||| | | | ||  |   |  |    ||       |
Db        433 TTCAAGGCCCTGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCC 492

Qy        559 ATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCC 618
              |||  |  |  || |||| |       | || || | ||   |  | ||   ||  |  |
Db        493 GCCCTCAAGGGCATGTGCGACGGCGGCGTCAACGTCCCC---CACAGCGAGACCCGCTTC 549

Qy        619 GTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATC 678
              ||   | ||| ||||  ||| | |  |  | |    |   |  ||  || | |  | ||
Db        550 GTCGGCTTCAACGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTC 609

Qy        679 AGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAG 738
              ||  |  ||   ||| |  | |||||||       |    ||||| | ||    |  |
Db        610 GGCGGCCACGTCGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGAC 669

Qy        739 CAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAG 798
              |    |  |  |  | | |  |||       | |    ||| |||   |||| ||||  |
Db        670 CGCCAGTTCTCGCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTAC 729

Qy        799 AAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCC 858
              |  |||| | | |     |    || |  |  || |  |  |  ||  | ||  |
Db        730 ACCGAGGCCCACAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAG 789

Qy        859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAG 918
              | ||        |  ||   | ||| ||||||    |        | |   | ||||
Db        790 CCCGAGGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAG 849

Qy        919 CGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
              |  |   | |     | |||| |  ||||||  |||| || || ||
Db        850 AACCGCGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCC 897
```

Search completed: December 31, 2005, 02:31:26
Job time : 8703.67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 11926.2 Seconds
(without alignments)
11710.708 Million cell updates/sec

Title: US-09-610-313B-32
Perfect score: 2457
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2457	100.0	2457	6	AX455916	AX455916 Sequence
2	2436.2	99.2	2463	6	AX455915	AX455915 Sequence
3	2415.4	98.3	2469	6	AX455914	AX455914 Sequence
4	2040	83.0	2300	6	BD263705	BD263705 Improved
5	2040	83.0	2300	6	CQ870575	CQ870575 Sequence
6	2040	83.0	2300	6	AR373388	AR373388 Sequence
7	2019.2	82.2	2306	6	BD263704	BD263704 Improved
8	2019.2	82.2	2306	6	CQ870574	CQ870574 Sequence
9	2019.2	82.2	2306	6	AR373387	AR373387 Sequence
10	2000.8	81.4	9166	6	AX427930	AX427930 Sequence
11	1998.4	81.3	2312	6	BD263706	BD263706 Improved
12	1998.4	81.3	2312	6	CQ870576	CQ870576 Sequence
13	1998.4	81.3	2312	6	AR373389	AR373389 Sequence
14	1992.4	81.1	9788	6	AX427936	AX427936 Sequence
15	1978.4	80.5	9169	6	AX427931	AX427931 Sequence
16	1975	80.4	9194	6	AX427926	AX427926 Sequence
17	1973.4	80.3	9194	6	AX427925	AX427925 Sequence
18	1973.4	80.3	12411	6	AX427927	AX427927 Sequence
19	1970	80.2	9785	6	AX427938	AX427938 Sequence
20	1966.6	80.0	9167	6	AX427933	AX427933 Sequence
21	1966.6	80.0	9170	6	AX427928	AX427928 Sequence
22	1966.6	80.0	9782	6	AX427935	AX427935 Sequence
23	1966.6	80.0	9783	6	AX427934	AX427934 Sequence
24	1966.6	80.0	9790	6	CS078943	CS078943 Sequence
25	1966.6	80.0	9792	6	AX427932	AX427932 Sequence
c 26	1966.6	80.0	36066	6	CS078956	CS078956 Sequence
27	1965	80.0	9189	6	AX427921	AX427921 Sequence
28	1962	79.9	4353	11	AF287353	AF287353 Synthetic
29	1945.4	79.2	3015	6	AX455946	AX455946 Sequence
30	1928.8	78.5	3009	6	AX455987	AX455987 Sequence
31	1915.8	78.0	4319	6	BD263639	BD263639 Improved
32	1915.8	78.0	4319	6	CQ870498	CQ870498 Sequence
33	1915.8	78.0	4319	6	AR373322	AR373322 Sequence
34	1905.2	77.5	4352	11	AF287352	AF287352 Synthetic
35	1905.2	77.5	4352	11	AY789509	AY789509 Synthetic
36	1878.6	76.5	1965	6	AX455952	AX455952 Sequence
37	1872.8	76.2	2299	6	BD263703	BD263703 Improved
38	1872.8	76.2	2299	6	CQ870573	CQ870573 Sequence
39	1872.8	76.2	2299	6	AR373386	AR373386 Sequence
40	1859.8	75.7	2577	6	AX457088	AX457088 Sequence
41	1852.2	75.4	4343	6	AX306428	AX306428 Sequence
42	1852	75.4	2305	6	BD263702	BD263702 Improved
43	1852	75.4	2305	6	CQ870572	CQ870572 Sequence
44	1852	75.4	2305	6	AR373385	AR373385 Sequence
45	1848.2	75.2	4341	6	AX306429	AX306429 Sequence

ALIGNMENTS

RESULT 1
AX455916
LOCUS AX455916 2457 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 32 from Patent WO0204493.
ACCESSION AX455916
VERSION AX455916.1 GI:21714901
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof
JOURNAL Patent: WO 0204493-A 32 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES Location/Qualifiers
source 1. .2457
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YMWM"
ORIGIN

Query Match 100.0%; Score 2457; DB 6; Length 2457;
Best Local Similarity 100.0%; Pred. No. 8.9e-203;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
```

```
Db      181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Qy      241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db      241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Qy      301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db      301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Qy      361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db      361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Qy      421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db      421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Qy      481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db      481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Qy      541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db      541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Qy      601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db      601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Qy      661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db      661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Qy      721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db      721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Qy      781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db      781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Qy      841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db      841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy      901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db      901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy      961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db      961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy     1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db     1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy     1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db     1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy     1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db     1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy     1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db     1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Qy     1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db     1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy     1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380
Db     1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380
Qy     1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
Db     1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
Qy     1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
Db     1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
Qy     1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
Db     1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
Qy     1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
Db     1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
Qy     1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
Db     1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
Qy     1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
Db     1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
Qy     1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
Db     1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
Qy     1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
Db     1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
Qy     1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
Db     1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
Qy     1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
Db     1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
Qy     1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
Db     1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
Qy     2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
Db     2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
Qy     2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
Db     2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
Qy     2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
Db     2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
Qy     2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
Db     2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
Qy     2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
Db     2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
Qy     2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400
Db     2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400
```

```
Qy       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457


RESULT 2
AX455915
LOCUS       AX455915               2463 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 31 from Patent WO0204493.
ACCESSION   AX455915
VERSION     AX455915.1  GI:21714900
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 31 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES             Location/Qualifiers
     source          1. .2463
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975YM"
ORIGIN

  Query Match             99.2%;  Score 2436.2;  DB 6;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 5.5e-201;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              |||||||||||||||||||||||||||||||||||||    ||      ||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
```

```
Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2455 TTC 2457
             |||
Db       2461 TTC 2463
```

```
RESULT 3
AX455914
LOCUS       AX455914               2469 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 30 from Patent WO0204493.
ACCESSION   AX455914
VERSION     AX455914.1  GI:21714899
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
  TITLE     Polynucleotides encoding antigenic hiv type c polypeptides,
            polypeptides and uses thereof
  JOURNAL   Patent: WO 0204493-A 30 17-JAN-2002;
            CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
```

```
FEATURES             Location/Qualifiers
     source          1. .2469
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PR975(+)"
ORIGIN

  Query Match             98.3%;  Score 2415.4;  DB 6;  Length 2469;
  Best Local Similarity   99.3%;  Pred. No. 3.3e-199;
  Matches 2451;  Conservative    0;  Mismatches    6;  Indels   12;  Gaps    2;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
```

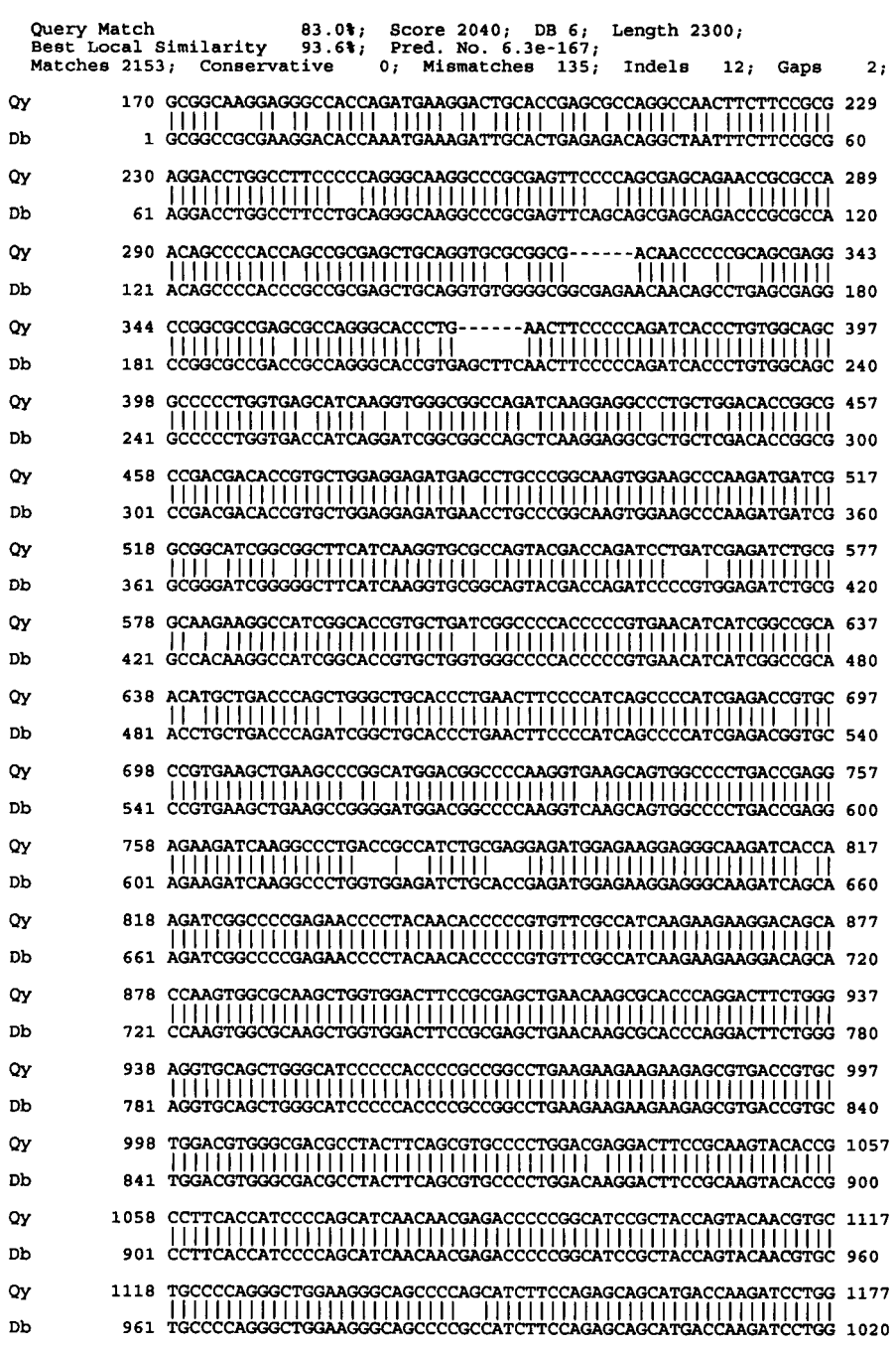
```
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Qy       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Qy       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Qy       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Qy       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Qy       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Qy       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Qy       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Qy       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Qy       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Qy       1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Qy       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
```

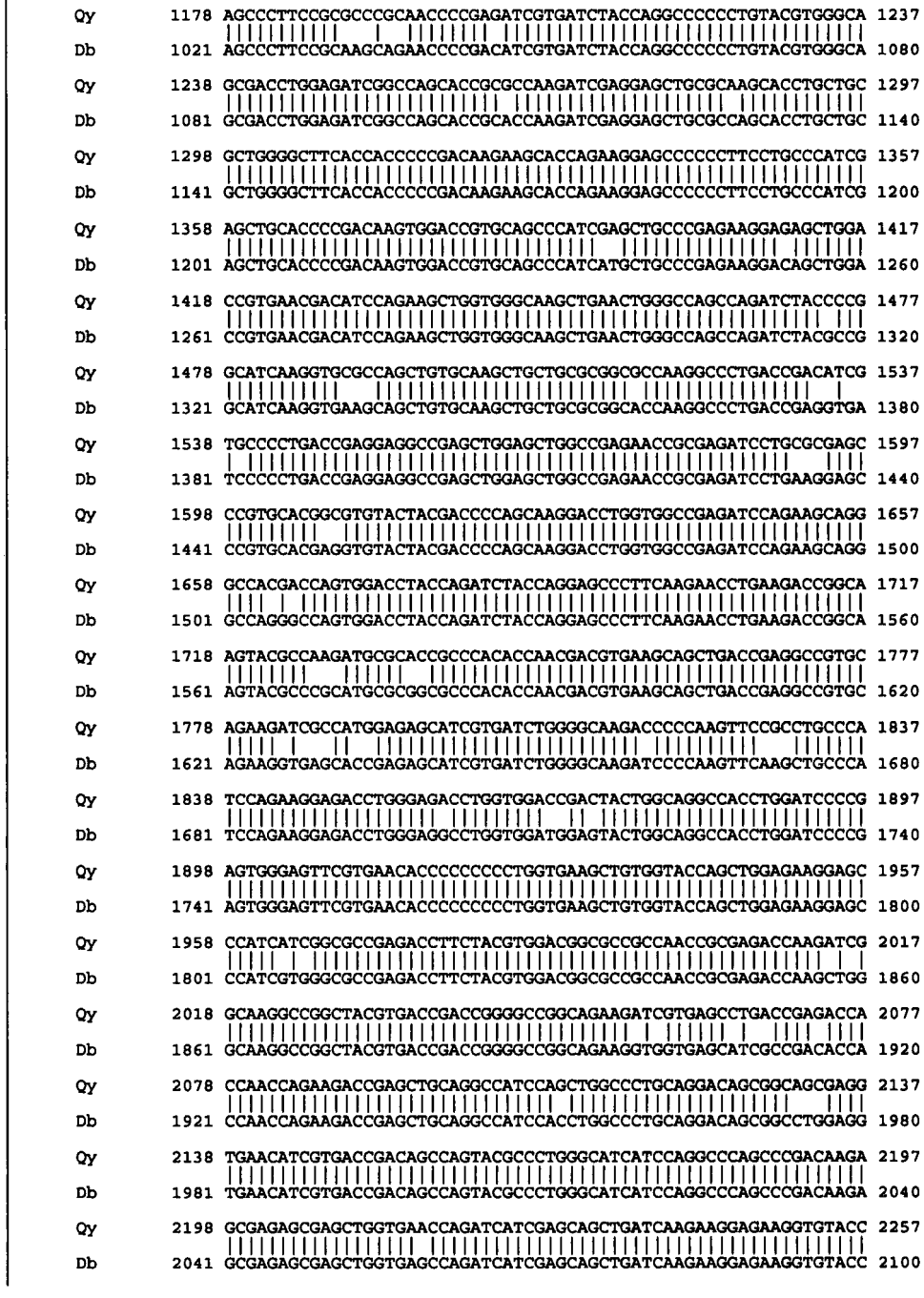
```
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Qy       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Qy       2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Qy       2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Qy       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Qy       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Qy       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Qy       2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Qy       2449 GGTGAATTC 2457
Db       2461 GGTGAATTC 2469
```

```
RESULT 4
BD263705
LOCUS       BD263705               2300 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263705
VERSION     BD263705.1  GI:33073473
KEYWORDS    JP 2002533124-A/72.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2300)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 72 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/72
            PD   08-OCT-2002
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt.YMWM FH
            Key             Location/Qualifiers
            FT   source          1. .2300
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .2300
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
```

ORIGIN

Query Match 83.0%; Score 2040; DB 6; Length 2300;
Best Local Similarity 93.6%; Pred. No. 6.3e-167;
Matches 2153; Conservative 0; Mismatches 135; Indels 12; Gaps 2;





```
Qy       2258 TGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317
Db       2101 TGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA 2160

Qy       2318 GCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACC 2377
Db       2161 GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC 2220

Qy       2378 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2437
Db       2221 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2280

Qy       2438 GGGCTAGCACCGGTGAATTC 2457
Db       2281 GGGCTAGCACCGGTGAATTC 2300


RESULT 5
CQ870575
LOCUS       CQ870575               2300 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 83 from Patent EP1433851.
ACCESSION   CQ870575
VERSION     CQ870575.1  GI:52000091
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 83 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2300
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt.YMWM"
ORIGIN

  Query Match             83.0%;  Score 2040;  DB 6;  Length 2300;
  Best Local Similarity   93.6%;  Pred. No. 6.3e-167;
  Matches 2153;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1357
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1200

Qy       1358 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGA 1417
Db       1201 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGA 1260

Qy       1418 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1477
Db       1261 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG 1320

Qy       1478 GCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG 1537
Db       1321 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGA 1380

Qy       1538 TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGC 1597
Db       1381 TCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC 1440

Qy       1598 CCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1657
Db       1441 CCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1500
```

```
Qy       1658 GCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1717
Db       1501 GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1560

Qy       1718 AGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1777
Db       1561 AGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1620

Qy       1778 AGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCA 1837
Db       1621 AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA 1680

Qy       1838 TCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCG 1897
Db       1681 TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG 1740

Qy       1898 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1957
Db       1741 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1800

Qy       1958 CCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCG 2017
Db       1801 CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG 1860

Qy       2018 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCA 2077
Db       1861 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA 1920

Qy       2078 CCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGG 2137
Db       1921 CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG 1980

Qy       2138 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2197
Db       1981 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2040

Qy       2198 GCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2257
Db       2041 GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2100

Qy       2258 TGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317
Db       2101 TGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA 2160

Qy       2318 GCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACC 2377
Db       2161 GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC 2220

Qy       2378 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2437
Db       2221 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2280

Qy       2438 GGGCTAGCACCGGTGAATTC 2457
Db       2281 GGGCTAGCACCGGTGAATTC 2300
```

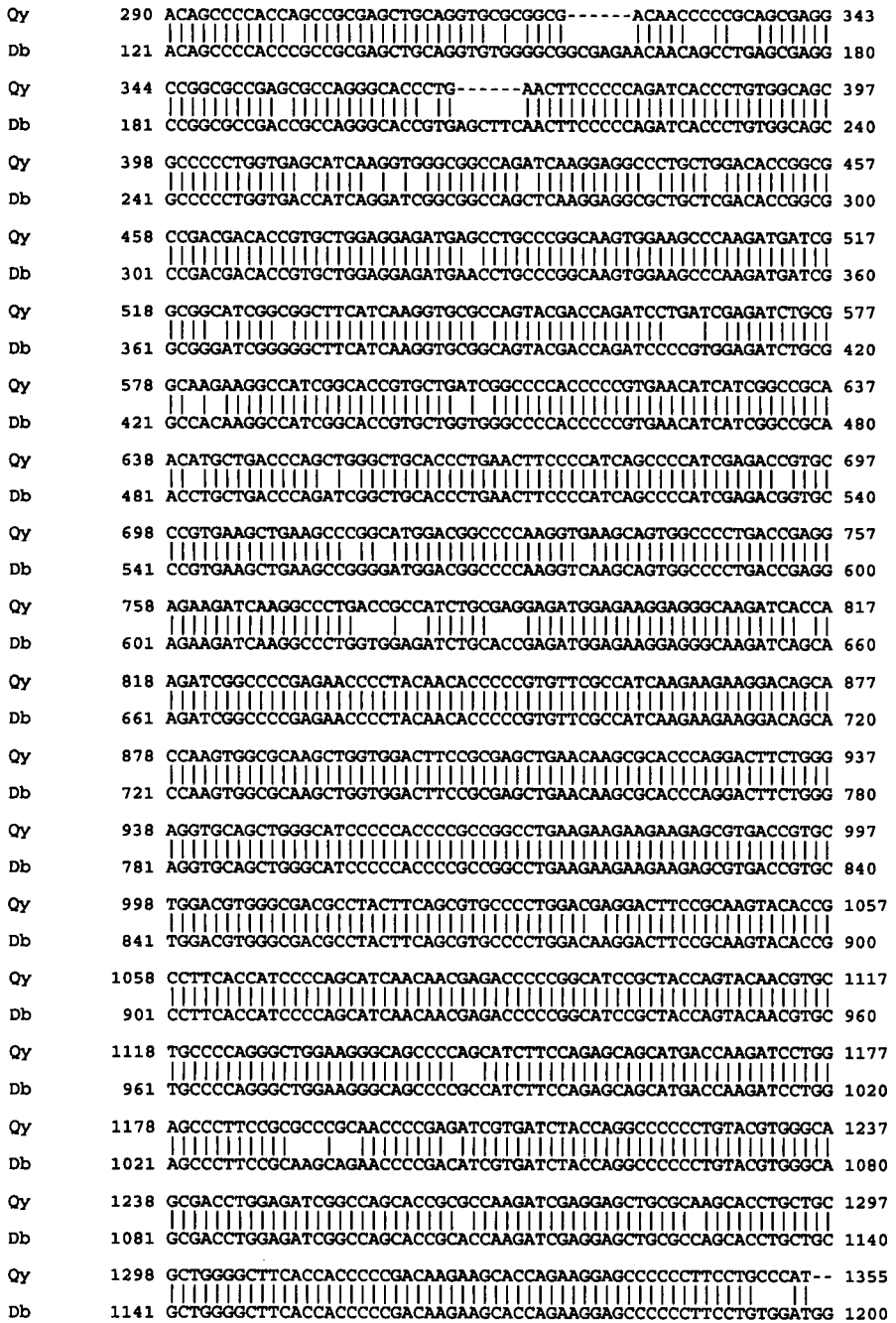
```
RESULT 6
AR373388
LOCUS       AR373388               2300 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 83 from patent US 6602705.
ACCESSION   AR373388
VERSION     AR373388.1  GI:40075491
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2300)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 83 05-AUG-2003;
            Chiron Corporation; Emeryville, CA
FEATURES             Location/Qualifiers
     source          1. .2300
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             83.0%;  Score 2040;  DB 6;  Length 2300;
  Best Local Similarity   93.6%;  Pred. No. 6.3e-167;
  Matches 2153;  Conservative    0;  Mismatches  135;  Indels   12;  Gaps    2;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
```

```
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297

Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1357

Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCG 1200

Qy       1358 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGA 1417

Db       1201 AGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGA 1260

Qy       1418 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1477

Db       1261 CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG 1320

Qy       1478 GCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCG 1537

Db       1321 GCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGA 1380

Qy       1538 TGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGC 1597

Db       1381 TCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC 1440

Qy       1598 CCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1657

Db       1441 CCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG 1500

Qy       1658 GCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1717

Db       1501 GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA 1560

Qy       1718 AGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1777

Db       1561 AGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC 1620

Qy       1778 AGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCA 1837

Db       1621 AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA 1680

Qy       1838 TCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCG 1897

Db       1681 TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG 1740

Qy       1898 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1957

Db       1741 AGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC 1800

Qy       1958 CCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCG 2017

Db       1801 CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGG 1860

Qy       2018 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCA 2077

Db       1861 GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA 1920

Qy       2078 CCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGG 2137

Db       1921 CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG 1980

Qy       2138 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2197

Db       1981 TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGA 2040

Qy       2198 GCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2257

Db       2041 GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC 2100

Qy       2258 TGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317

Db       2101 TGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA 2160

Qy       2318 GCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACC 2377

Db       2161 GCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGATCTACC 2220

Qy       2378 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2437

Db       2221 AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCG 2280

Qy       2438 GGGCTAGCACCGGTGAATTC 2457

Db       2281 GGGCTAGCACCGGTGAATTC 2300
```

RESULT 7
BD263704
LOCUS BD263704 2306 bp DNA linear PAT 17-JUL-2003
DEFINITION Improved expression of HIV polypeptides and production of virus-like particles.
ACCESSION BD263704
VERSION BD263704.1 GI:33073472
KEYWORDS JP 2002533124-A/71.
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 2306)
AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.
TITLE Improved expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: JP 2002533124-A 71 08-OCT-2002;
CHIRON CORP
COMMENT OS Artificial Sequence
PN JP 2002533124-A/71
PD 08-OCT-2002
PF 30-DEC-1999 JP 2000591193
PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI
SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
KARIN HARTOG,
PI HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
C12N5/10,
PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH Key
Location/Qualifiers
FT source 1. .2306
FT /organism='Artificial Sequence'.
FEATURES Location/Qualifiers
source 1. .2306
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN

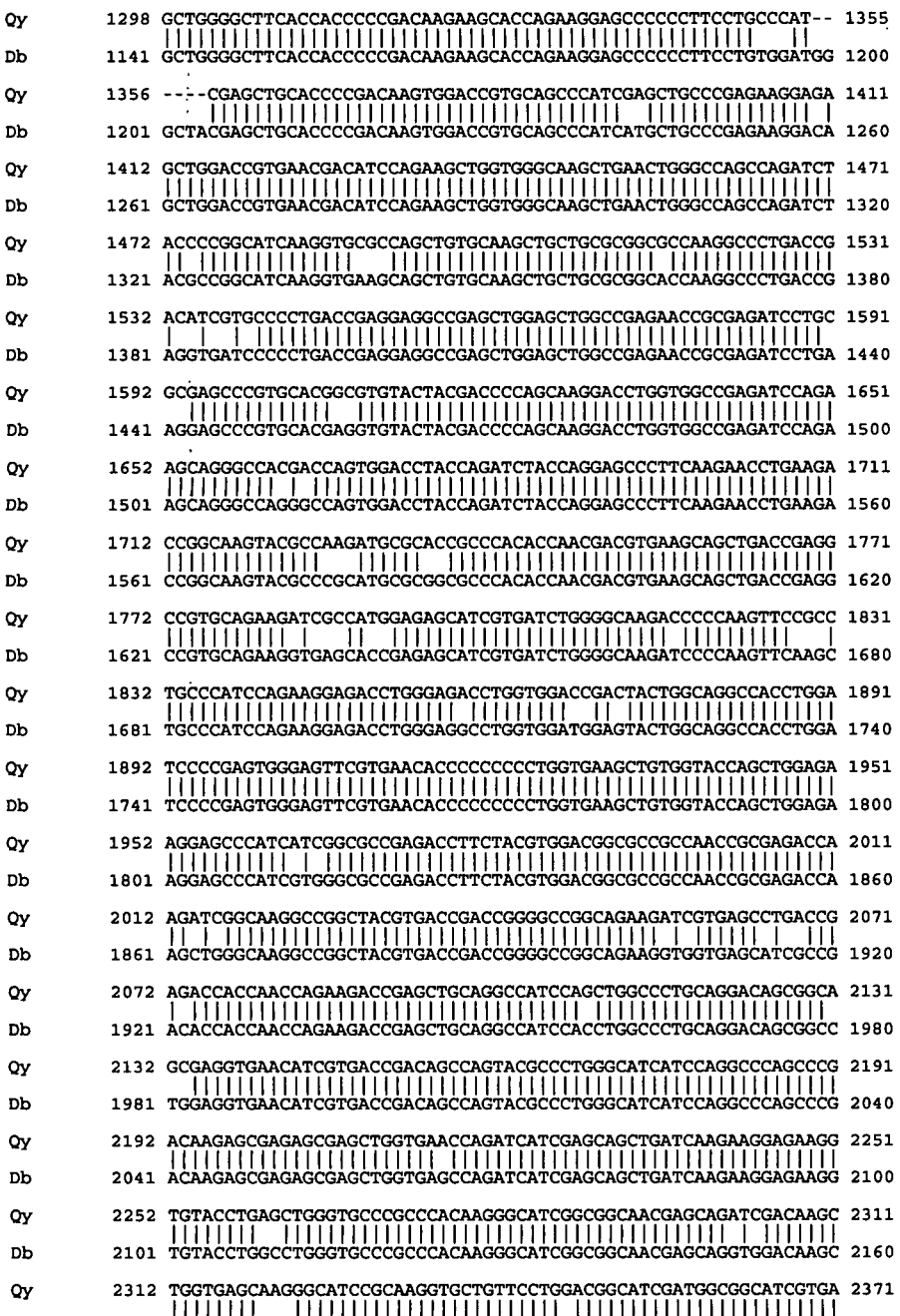
Query Match 82.2%; Score 2019.2; DB 6; Length 2306;
Best Local Similarity 93.2%; Pred. No. 3.9e-165;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

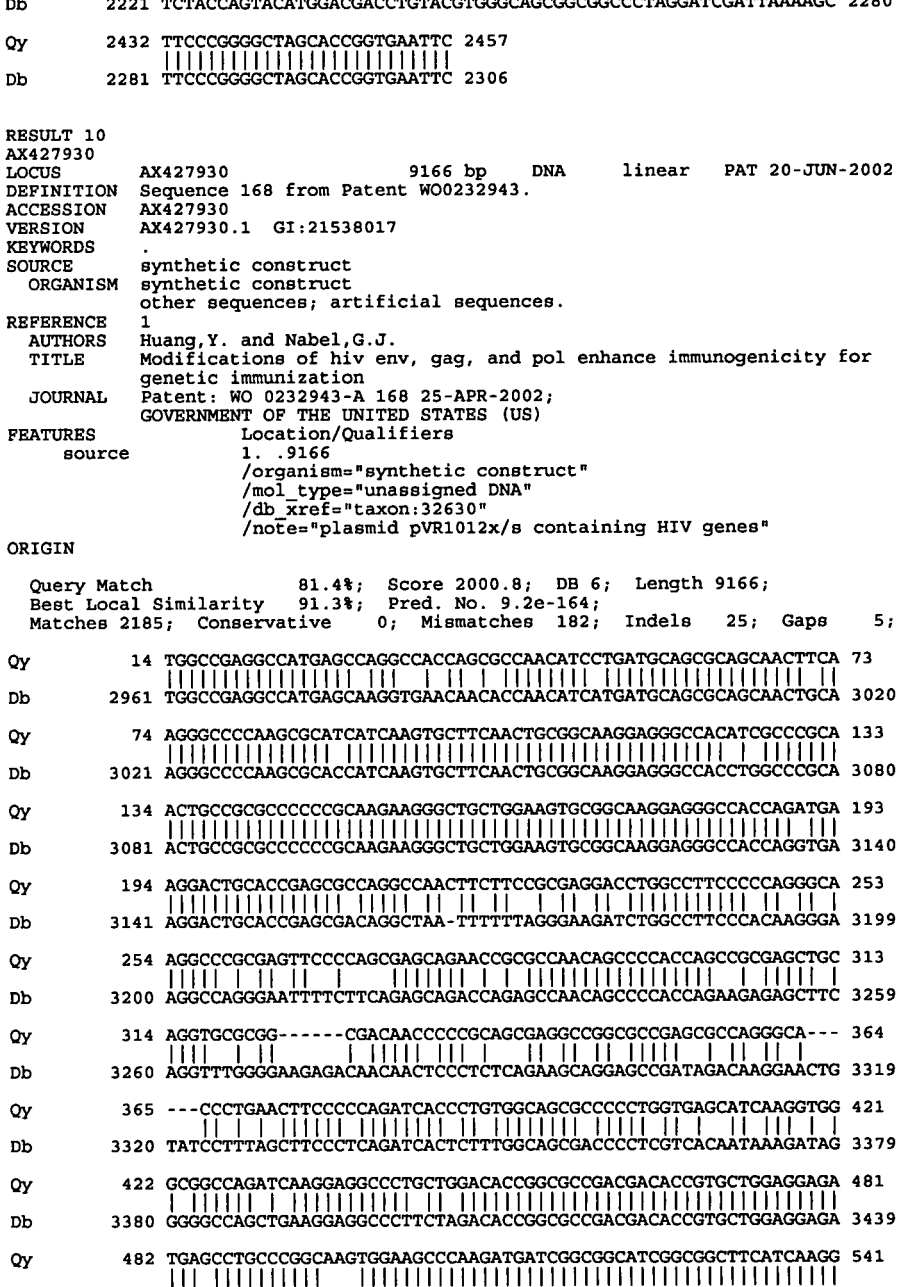
```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120
```


Qy 290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db 121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180
Qy 344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db 181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240
Qy 398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db 241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300
Qy 458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db 301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360
Qy 518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db 361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420
Qy 578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db 421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480
Qy 638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db 481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540
Qy 698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db 541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600
Qy 758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db 601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660
Qy 818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db 661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720
Qy 878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db 721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780
Qy 938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db 781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840
Qy 998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db 841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900
Qy 1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db 901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960
Qy 1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db 961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020
Qy 1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db 1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080
Qy 1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db 1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140
Qy 1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1355
Db 1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200
Qy 1356 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1411
Db 1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260
Qy 1412 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1471
Db 1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320
Qy 1472 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1531
Db 1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380
Qy 1532 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1591
Db 1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440
Qy 1592 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1651
Db 1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500
Qy 1652 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1711
Db 1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560
Qy 1712 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1771
Db 1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620
Qy 1772 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1831
Db 1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680
Qy 1832 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1891
Db 1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740
Qy 1892 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1951
Db 1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800
Qy 1952 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2011
Db 1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860
Qy 2012 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2071
Db 1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920
Qy 2072 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2131
Db 1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980
Qy 2132 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2191
Db 1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040
Qy 2192 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2251
Db 2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100
Qy 2252 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2311
Db 2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160
Qy 2312 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2371
Db 2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220
Qy 2372 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2431
Db 2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280
Qy 2432 TTCCCGGGGCTAGCACCGGTGAATTC 2457

```
Db       2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306

RESULT 8
CQ870574
LOCUS       CQ870574                2306 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 82 from Patent EP1433851.
ACCESSION   CQ870574
VERSION     CQ870574.1  GI:52000090
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
            Liu,H., Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: EP 1433851-A 82 30-JUN-2004;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..2306
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Description of Artificial Sequence:
                     FS(-).protmod.RTopt.YM"
ORIGIN

  Query Match             82.2%;  Score 2019.2;  DB 6;  Length 2306;
  Best Local Similarity   93.2%;  Pred. No. 3.9e-165;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140

Qy       1298 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT-- 1355
Db       1141 GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGG 1200

Qy       1356 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGA 1411
Db       1201 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1260

Qy       1412 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1471
Db       1261 GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1320

Qy       1472 ACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCG 1531
Db       1321 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG 1380

Qy       1532 ACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGC 1591
Db       1381 AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1440

Qy       1592 GCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1651
Db       1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGA 1500

Qy       1652 AGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1711
Db       1501 AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA 1560

Qy       1712 CCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1771
Db       1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGG 1620

Qy       1772 CCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCC 1831
Db       1621 CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680
```

```
Qy       1832 TGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1891
             |||||||||||||||||||||||||| |||||||||  || ||||||||||||||||||||
Db       1681 TGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740

Qy       1892 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1951
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA 1800

Qy       1952 AGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 2011
             |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCA 1860

Qy       2012 AGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCG 2071
             || | ||||||||||||||||||||||||||||||||||| | |||||| |  |||
Db       1861 AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCG 1920

Qy       2072 AGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2131
             | ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1921 ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1980

Qy       2132 GCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCG 2040

Qy       2192 ACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2251
             |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       2041 ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2100

Qy       2252 TGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2311
             ||||||||  |||||||||||||||||||||||||||||||||||| | |||||||
Db       2101 TGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC 2160

Qy       2312 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGA 2371
             ||||||||   |||||||||||||||||||||| ||||||||||||||||||||||||||
Db       2161 TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGA 2220

Qy       2372 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2431
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGC 2280

Qy       2432 TTCCCCGGGGCTAGCACCGGTGAATTC 2457
             |||||||||||||||||||||||||||||
Db       2281 TTCCCCGGGGCTAGCACCGGTGAATTC 2306


RESULT 9
AR373387
LOCUS       AR373387               2306 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 82 from patent US 6602705.
ACCESSION   AR373387
VERSION     AR373387.1  GI:40075490
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2306)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 82 05-AUG-2003;
            Chiron Corporation; Emeryville, CA
FEATURES             Location/Qualifiers
     source          1. .2306
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           82.2%;  Score 2019.2;  DB 6;  Length 2306;
  Best Local Similarity   93.2%;  Pred. No. 3.9e-165;
  Matches 2150;  Conservative    0;  Mismatches  138;  Indels   18;  Gaps    3;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              |||||||||||||| |||||||||||||||||||| ||||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              |||||||||| |||||||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      |||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
              |||||||||| ||||| | | ||||||||| ||||||||| ||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
              |||| ||||| |||||||||||||||| ||||||||||||||    | ||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
              || | |||||||||||||||||| | ||||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
              || |||||||||| | ||||||||||||||||||||||||||||||||||||| ||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
              |||||||||||||| || |||||||||||||| ||||||||||||||||||||||||||
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
              |||||||||||||||    |  ||||||   |||||||||||||||||||||||||| ||
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
              |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1237
              |||||||||||  |||||||| ||||||||||||||||||||||||||||||||||||||
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA 1080

Qy       1238 GCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGC 1297
              |||||||||||||||||||||||||||||| |||||||||||||||||||| ||||||||
Db       1081 GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC 1140
```



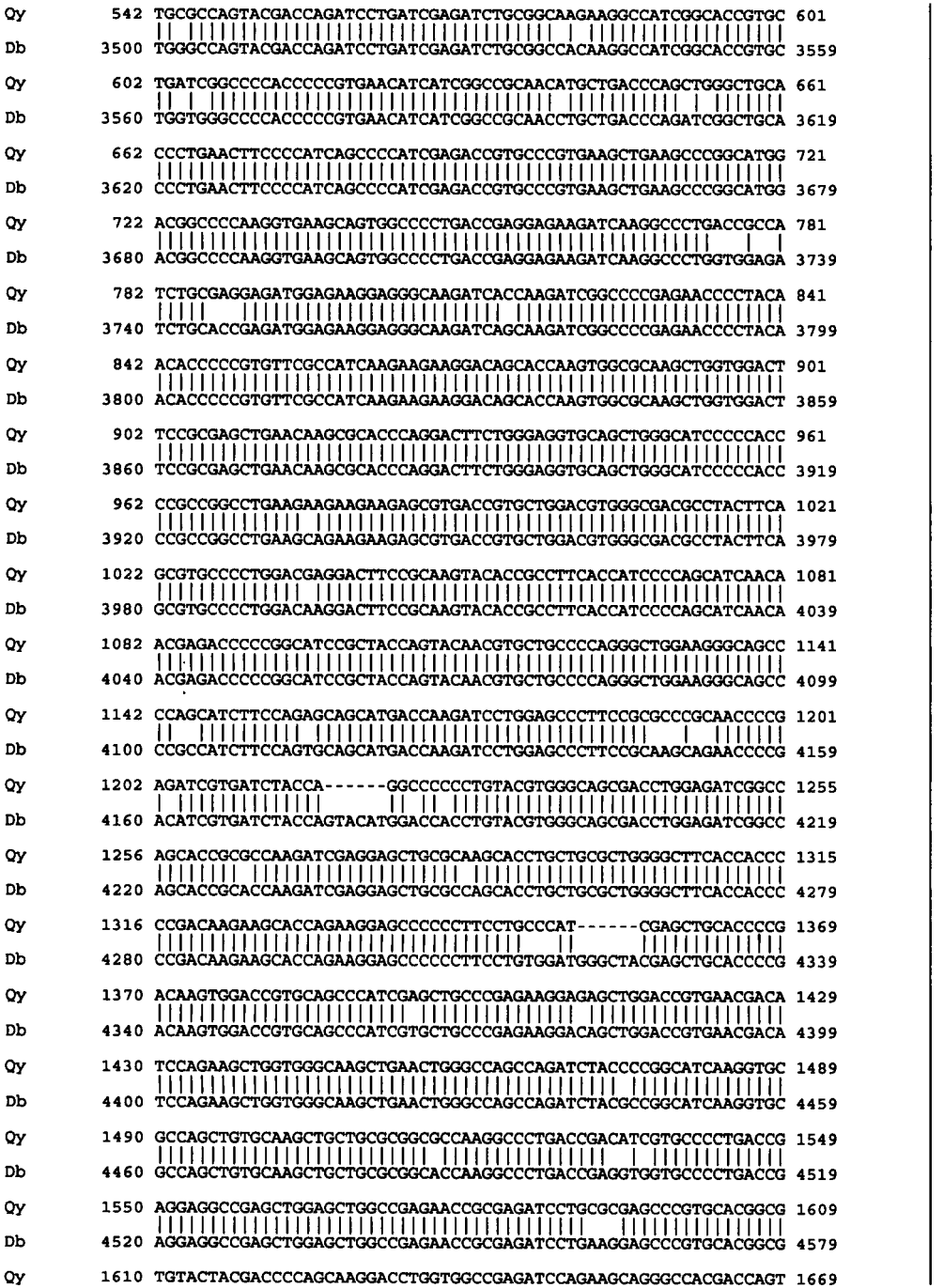


RESULT 10
AX427930
LOCUS AX427930 9166 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 168 from Patent WO0232943.
ACCESSION AX427930
VERSION AX427930.1 GI:21538017
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Huang,Y. and Nabel,G.J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization
JOURNAL Patent: WO 0232943-A 168 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)

FEATURES Location/Qualifiers
source 1. .9166
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match 81.4%; Score 2000.8; DB 6; Length 9166;
Best Local Similarity 91.3%; Pred. No. 9.2e-164;
Matches 2185; Conservative 0; Mismatches 182; Indels 25; Gaps 5;



```
Qy       542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
Db      3500 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3559

Qy       602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
Db      3560 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3619

Qy       662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
Db      3620 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3679

Qy       722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db      3680 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3739

Qy       782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
Db      3740 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3799

Qy       842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
Db      3800 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3859

Qy       902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
Db      3860 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3919

Qy       962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
Db      3920 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3979

Qy      1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
Db      3980 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4039

Qy      1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
Db      4040 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4099

Qy      1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
Db      4100 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4159

Qy      1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
Db      4160 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4219

Qy      1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
Db      4220 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4279

Qy      1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCG 1369
Db      4280 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4339

Qy      1370 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1429
Db      4340 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4399

Qy      1430 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1489
Db      4400 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4459

Qy      1490 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1549
Db      4460 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4519

Qy      1550 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1609
Db      4520 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4579

Qy      1610 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1669
Db      4580 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4639

Qy      1670 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1729
Db      4640 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4699

Qy      1730 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1789
Db      4700 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4759

Qy      1790 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1849
Db      4760 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4819

Qy      1850 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1909
Db      4820 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4879

Qy      1910 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1969
Db      4880 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4939

Qy      1970 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2029
Db      4940 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 4999

Qy      2030 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2089
Db      5000 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5059

Qy      2090 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2149
Db      5060 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5119

Qy      2150 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2209
Db      5120 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5179

Qy      2210 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2269
Db      5180 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5239

Qy      2270 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2329
Db      5240 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5299

Qy      2330 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381
Db      5300 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5351
```

```
RESULT 11
BD263706
LOCUS       BD263706               2312 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Improved expression of HIV polypeptides and production of
            virus-like particles.
ACCESSION   BD263706
VERSION     BD263706.1  GI:33073474
KEYWORDS    JP 2002533124-A/73.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
            Greer,C., Selby,M. and Walker,C.
  TITLE     Improved expression of HIV polypeptides and production of
            virus-like particles
  JOURNAL   Patent: JP 2002533124-A 73 08-OCT-2002;
            CHIRON CORP
COMMENT     OS   Artificial Sequence
            PN   JP 2002533124-A/73
            PD   08-OCT-2002
```

```
            PF   30-DEC-1999 JP 2000591193
            PR   31-DEC-1998 US    60/114495,01-DEC-1999 US    60/168471 PI
            SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI
            KARIN HARTOG,
            PI   HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
            C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
            C12N5/10,
            PC   C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
            Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH   Key
                      Location/Qualifiers
            FT   source          1. .2312
            FT                   /organism='Artificial Sequence'.
FEATURES              Location/Qualifiers
     source           1. .2312
                      /organism="synthetic construct"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:32630"
ORIGIN

  Query Match             81.3%;  Score 1998.4;  DB 6;  Length 2312;
  Best Local Similarity   92.9%;  Pred. No. 2.4e-163;
  Matches 2147;  Conservative    0;  Mismatches  141;  Indels   24;  Gaps    4;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              |||||||||||||||  |||||||||||||||||||||  |||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              |||||||||| |||||||||||||||||| | ||||      |||||  ||  ||||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      |||||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
              |||||||||||| ||||| | | ||||||||| ||||||||| ||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
              |||| ||||| |||||||||||||||| ||||||||||||||||    | ||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
              || | |||||||||||||||||||| | ||||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
              || |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
              |||||||||||||||| || |||||||||||||||| ||||||||||||||||||||||||
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
              ||||||||||||||||    | ||||||   |||||||||||||||||||||||||| ||
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
              |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
              |||||||||||  | |||||||| ||||||||||||||||      || |  ||||||||
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
              |||||||||||||||||||||||||||||||| |||||||||||||||||| ||||||
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGC 1351
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 CCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1405
                ||      |||||||||||||||||||||||||||||||||||||||  ||||||||||
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1406 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1465
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1466 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1525
              |||||||| |||||||||||||   ||||||||||||||||||||||| |||||||||
Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1526 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1585
              |||||||  |  | ||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1586 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1645
              |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1646 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1705
              |||||||||||||||  | ||||||||||||||||||||||||||||||||||| ||||
Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1706 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1765
              ||||||||||||||||||||  ||||||  ||||||||||||||||||||||||||||||
Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1766 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1825
              |||||||||||||||||  ||  |||||||||||||||||||||||||||||| ||||||||
Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1826 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1885
              ||  |||||||||||||||||||||||||||||| |||||||||  || ||||||||||||
Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1886 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1945
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1946 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2005
              |||||||||||||||||  |||||||||||||||||||||||||||||||||| ||||||
Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860
```

```
Qy       2006 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2065
             |||||||| | ||||||||||||||||||||||||||||||||||||||| | ||||||
Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2066 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2125
             |  |||| |||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2126 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2185
             |||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2186 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2245
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2246 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2305
             |||||||||||||||  |||||||||||||||||||||||||||||||||||||||| | |
Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2306 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2365
             ||||||||||||||    |||||||||||||||||||||||| ||||||||||||||||||
Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2366 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2425
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2426 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
             |||||||||||||||||||||||||||||||
Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

RESULT 12
CQ870576
LOCUS CQ870576 2312 bp DNA linear PAT 13-SEP-2004
DEFINITION Sequence 84 from Patent EP1433851.
ACCESSION CQ870576
VERSION CQ870576.1 GI:52000092
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.
TITLE Improved expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: EP 1433851-A 84 30-JUN-2004;
CHIRON CORPORATION (US)
FEATURES Location/Qualifiers
source 1. .2312
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Description of Artificial Sequence:
FS(-).protmod.RTopt(+)"
ORIGIN

Query Match 81.3%; Score 1998.4; DB 6; Length 2312;
Best Local Similarity 92.9%; Pred. No. 2.4e-163;
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

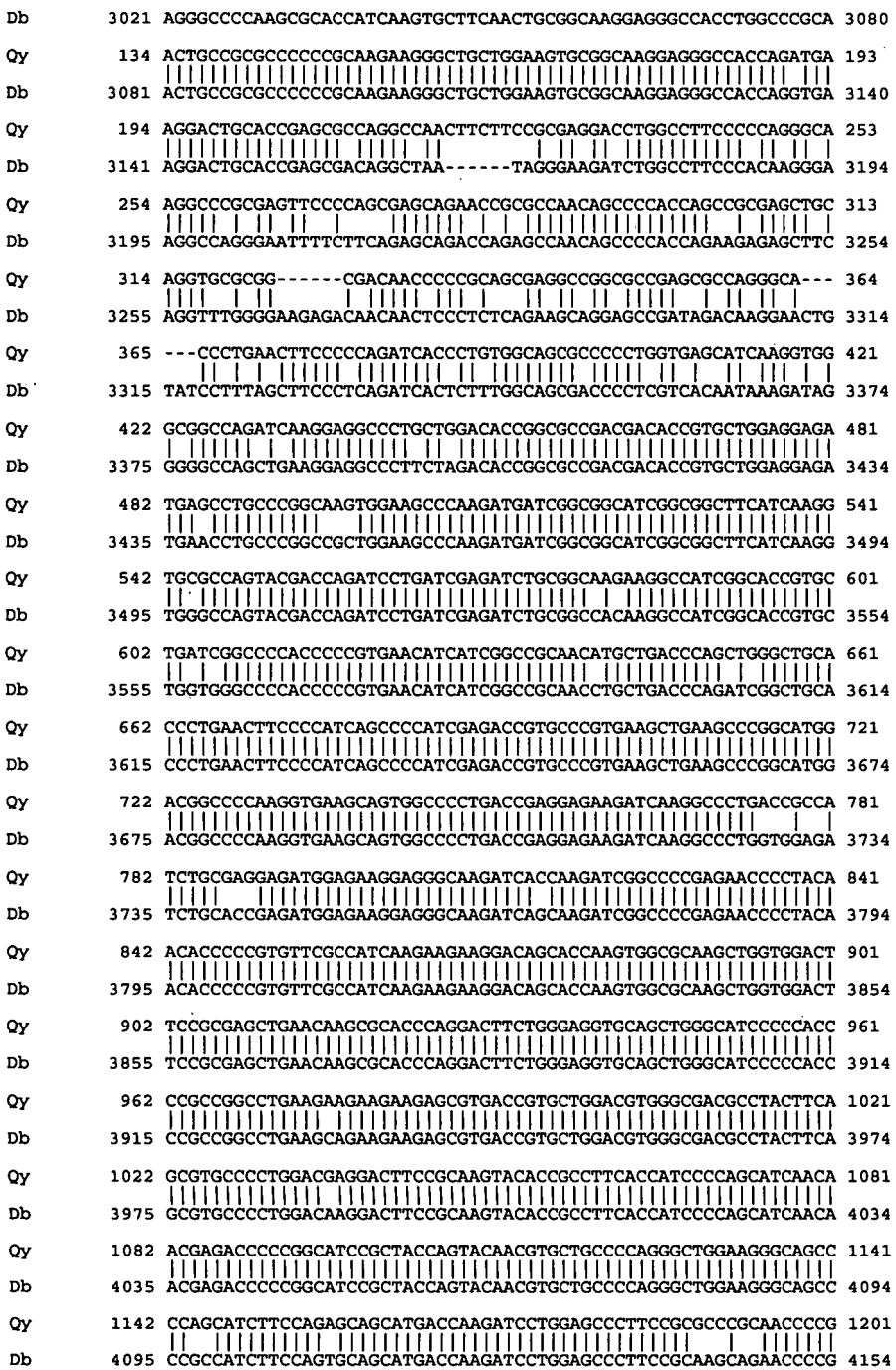
```
Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229
              |||||   || || ||||| ||||| || ||||| ||| | ||||| || ||||||||||
Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
              ||||||||||||||  |||||||||||||||||||||  |||||||||| ||||||||
Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343
              ||||||||||| |||||||||||||||||||||| | ||||      |||||  ||  |||||||
Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397
              |||||||||| |||||||||||| ||      |||||||||||||||||||||||||||||
Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457
              |||||||||||| ||||| | | ||||||||| |||||||||| ||||| ||||||||||
Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577
              |||| ||||| |||||||||||||||| ||||||||||||||    | ||||||||||
Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637
              || | |||||||||||||||||||| | ||||||||||||||||||||||||||||||||
Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697
              || |||||||||| | ||||||||||||||||||||||||||||||||||||| ||||
Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757
              |||||||||||||||||| || |||||||||||||||||| ||||||||||||||||||
Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
              |||||||||||||||||    |  ||||||  |||||||||||||||||||||||||| ||
Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
              |||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231
              |||||||||||   | |||||||| |||||||||||||      || | ||||||||
Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGC 1351
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 CCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1405
                ||      |||||||||||||||||||||||||||||||||||||  ||||||||||||
```

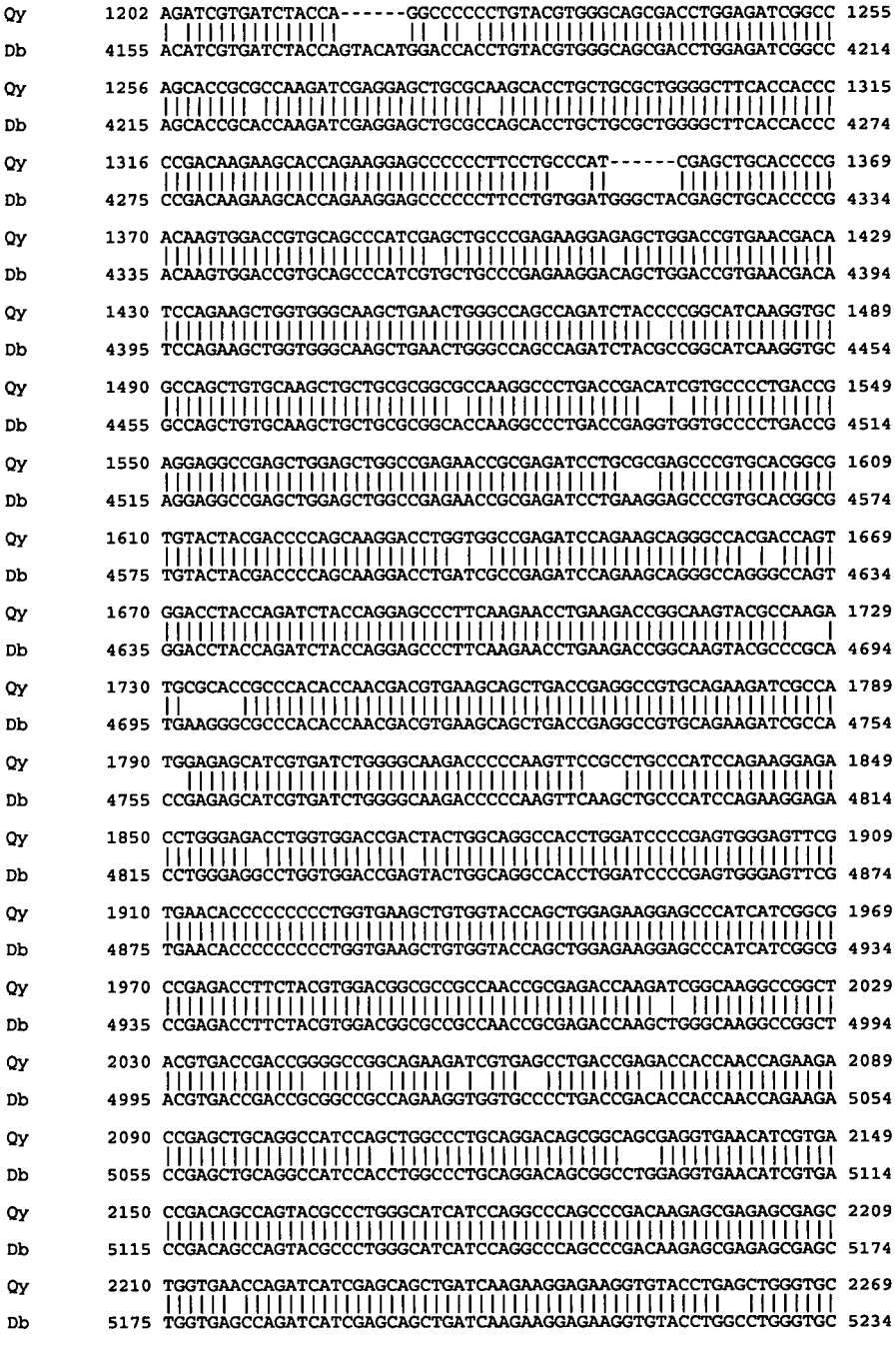
```
Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1406 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1465

Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1466 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1525

Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1526 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1585

Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1586 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1645

Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1646 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1705

Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1706 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1765

Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1766 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1825

Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1826 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1885

Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1886 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1945

Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1946 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2005

Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2006 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2065

Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2066 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2125

Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2126 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2185

Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2186 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2245

Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2246 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2305

Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2306 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2365

Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2366 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2425

Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2426 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457

Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 13
AR373389
LOCUS       AR373389               2312 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 84 from patent US 6602705.
ACCESSION   AR373389
VERSION     AR373389.1  GI:40075492
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Barnett,S.W., Megede,J., Greer,C. and Selby,M.
  TITLE     Expression of HIV polypeptides and production of virus-like
            particles
  JOURNAL   Patent: US 6602705-A 84 05-AUG-2003;
            Chiron Corporation; Emeryville, CA
FEATURES             Location/Qualifiers
     source          1. .2312
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             81.3%;  Score 1998.4;  DB 6;  Length 2312;
  Best Local Similarity   92.9%;  Pred. No. 2.4e-163;
  Matches 2147;  Conservative    0;  Mismatches  141;  Indels   24;  Gaps    4;

Qy        170 GCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCG 229

Db          1 GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

Qy        230 AGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289

Db         61 AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA 120

Qy        290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCG------ACAACCCCCGCAGCGAGG 343

Db        121 ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG 180

Qy        344 CCGGCGCCGAGCGCCAGGGCACCCTG------AACTTCCCCCAGATCACCCTGTGGCAGC 397

Db        181 CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240

Qy        398 GCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457

Db        241 GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG 300

Qy        458 CCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 517

Db        301 CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG 360

Qy        518 GCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCG 577

Db        361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG 420

Qy        578 GCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCA 637

Db        421 GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA 480

Qy        638 ACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGC 697

Db        481 ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC 540

Qy        698 CCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGG 757

Db        541 CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG 600

Qy        758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817

Db        601 AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660

Qy        818 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 877
```

```
Db        661 AGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCA 720

Qy        878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937

Db        721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780

Qy        938 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAAGCGTGACCGTGC 997

Db        781 AGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAAGCGTGACCGTGC 840

Qy        998 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCG 1057

Db        841 TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG 900

Qy       1058 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 1117

Db        901 CCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC 960

Qy       1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177

Db        961 TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020

Qy       1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA------GGCCCCCCTGTACG 1231

Db       1021 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG 1080

Qy       1232 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACC 1291

Db       1081 TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC 1140

Qy       1292 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGC 1351

Db       1141 TGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGT 1200

Qy       1352 CCAT------CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1405

Db       1201 GGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260

Qy       1406 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1465

Db       1261 AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC 1320

Qy       1466 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCC 1525

Db       1321 AGATCTACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC 1380

Qy       1526 TGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1585

Db       1381 TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440

Qy       1586 TCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1645

Db       1441 TCCTGAAGGAGCCCGTGCACGAGGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGA 1500

Qy       1646 TCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1705

Db       1501 TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC 1560

Qy       1706 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGA 1765

Db       1561 TGAAGACCGGCAAGTACGCCCGCATGCGCGGCGCCCACACCAACGACGTGAAGCAGCTGA 1620

Qy       1766 CCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGT 1825

Db       1621 CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680

Qy       1826 TCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1885

Db       1681 TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA 1740

Qy       1886 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1945

Db       1741 CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGC 1800

Qy       1946 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 2005

Db       1801 TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG 1860

Qy       2006 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCC 2065

Db       1861 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA 1920

Qy       2066 TGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2125

Db       1921 TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980

Qy       2126 GCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2185

Db       1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC 2040

Qy       2186 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGG 2245

Db       2041 AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100

Qy       2246 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCG 2305

Db       2101 AGAAGGTGTACCTGGCCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGG 2160

Qy       2306 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCA 2365

Db       2161 ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCA 2220

Qy       2366 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2425

Db       2221 TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT 2280

Qy       2426 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457

Db       2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 14
AX427936
LOCUS       AX427936               9788 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 174 from Patent WO0232943.
ACCESSION   AX427936
VERSION     AX427936.1  GI:21538023
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 174 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9788
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match            81.1%;  Score 1992.4;  DB 6;  Length 9788;
  Best Local Similarity   91.2%;  Pred. No. 4.8e-163;
  Matches 2181;  Conservative    0;  Mismatches  181;  Indels   30;  Gaps    5;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73

Db       2961 TGGCCGAGGCCATGAGCAAGGTGAACAACACCAACATCATGATGCAGCGCAGCAACTGCA 3020

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
```





```
Qy       2270 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2329
             ||||||||||||||||||||||||||||||||| | |||   |||||||||   |||||||
Db       5235 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5294

Qy       2330 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381
             |||||||||||||||||||||||||||||||    | |   | |     || || |
Db       5295 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5346


RESULT 15
AX427931
LOCUS       AX427931               9169 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 169 from Patent WO0232943.
ACCESSION   AX427931
VERSION     AX427931.1  GI:21538018
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Huang,Y. and Nabel,G.J.
  TITLE     Modifications of hiv env, gag, and pol enhance immunogenicity for
            genetic immunization
  JOURNAL   Patent: WO 0232943-A 169 25-APR-2002;
            GOVERNMENT OF THE UNITED STATES (US)
FEATURES             Location/Qualifiers
     source          1. .9169
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid pVR1012x/s containing HIV genes"
ORIGIN

  Query Match             80.5%;  Score 1978.4;  DB 6;  Length 9169;
  Best Local Similarity   91.1%;  Pred. No. 7.8e-162;
  Matches 2178;  Conservative    0;  Mismatches  186;  Indels   28;  Gaps    6;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||||||||||||||        | |||||||| |||||||||| |||||||||
Db       2967 TGGCCGAGGCCATGAGCCAGGTGCAGCACACCAACATCATGATGCAGCGCGGCAACTTCA 3026

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||  ||||||||   |||||||||||||||||||||||||| | |||||||
Db       3027 AGGGCCAGAAGCGCAT---CAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGGCCCGCA 3083

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db       3084 ACTGCCGCGCCCTGCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 3143

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              |||||||||||||||| ||||| || || ||   | || || |||||||||||| || || |
Db       3144 AGGACTGCACCGAGCGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGA 3202

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||| | || ||  |    ||||||| | | ||||||||||||||||  | |||||| |
Db       3203 AGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTC 3262

Qy        314 AGGTGCGCGG------CGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCA--- 364
              ||||  | ||       | ||||| ||| |   || || || |||||  | || || |
Db       3263 AGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAACTG 3322

Qy        365 ---CCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGG 421
                 || | | |||||| |||||||| || |||||||| ||||| || |  || ||| | |
Db       3323 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3382

Qy        422 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 481
              | ||||||  | |||||||||| || ||||||||||||||||||||||||||||||||
Db       3383 GGGGCCAGCTGAAGGAGGCCCTTCTAGACACCGGCGCCGACGACACCGTGCTGGAGGAGA 3442

Qy        482 TGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 541
              ||| ||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db       3443 TGAACCTGCCCGGCCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGG 3502

Qy        542 TGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGC 601
              || |||||||||||||||||||||||||||||||||||||||| | |||||||||||||||
Db       3503 TGGGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCCACAAGGCCATCGGCACCGTGC 3562

Qy        602 TGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCA 661
              || | |||||||||||||||||||||||||||||||||||| |||||||||| | |||||||
Db       3563 TGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCAACCTGCTGACCCAGATCGGCTGCA 3622

Qy        662 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3623 CCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGG 3682

Qy        722 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db       3683 ACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGGTGGAGA 3742

Qy        782 TCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACA 841
              |||||   |||||||||||||||||||||||| |||||||||||||||||||||||||||
Db       3743 TCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAACCCCTACA 3802

Qy        842 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3803 ACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3862

Qy        902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3863 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACC 3922

Qy        962 CCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 1021
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       3923 CCGCCGGCCTGAAGCAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCA 3982

Qy       1022 GCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 1081
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       3983 GCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACA 4042

Qy       1082 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4043 ACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCC 4102

Qy       1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201
              ||  |||||||||| ||||||||||||||||||||||||||||||    | |||||||
Db       4103 CCGCCATCTTCCAGTGCAGCATGACCAAGATCCTGGAGCCCTTCCGCAAGCAGAACCCCG 4162

Qy       1202 AGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
              | ||||||||||||||       || || ||||||||||||||||||||||||||||||
Db       4163 ACATCGTGATCTACCAGTACATGGACCACCTGTACGTGGGCAGCGACCTGGAGATCGGCC 4222

Qy       1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCC 1315
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4223 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCACCC 4282

Qy       1316 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCG 1369
              |||||||||||||||||||||||||||||||||||||||    ||      ||||||||||||
Db       4283 CCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCG 4342

Qy       1370 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACA 1429
              |||||||||||||||||||||||||| |||||||||||||| |||||||||||||||||||
Db       4343 ACAAGTGGACCGTGCAGCCCATCGTGCTGCCCGAGAAGGACAGCTGGACCGTGAACGACA 4402

Qy       1430 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGC 1489
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4403 TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGGCATCAAGGTGC 4462

Qy       1490 GCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCG 1549
              |||||||||||||||||||||||||||||| |||||||||||||| | ||||||||||||
Db       4463 GCCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGGTGCCCCTGACCG 4522

Qy       1550 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCG 1609
              |||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db       4523 AGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCCCGTGCACGGCG 4582

Qy       1610 TGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGT 1669
```

```
Db       4583 TGTACTACGACCCCAGCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGCCAGGGCCAGT 4642

Qy       1670 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA 1729
Db       4643 GGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGCA 4702

Qy       1730 TGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 1789
Db       4703 TGAAGGGCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCA 4762

Qy       1790 TGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGA 1849
Db       4763 CCGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGA 4822

Qy       1850 CCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 1909
Db       4823 CCTGGGAGGCCTGGTGGACCGAGTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCG 4882

Qy       1910 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 1969
Db       4883 TGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCG 4942

Qy       1970 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2029
Db       4943 CCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGGCT 5002

Qy       2030 ACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGA 2089
Db       5003 ACGTGACCGACCGCGGCCGCCAGAAGGTGGTGCCCCTGACCGACACCACCAACCAGAAGA 5062

Qy       2090 CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGA 2149
Db       5063 CCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGTGAACATCGTGA 5122

Qy       2150 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 2209
Db       5123 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGC 5182

Qy       2210 TGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGC 2269
Db       5183 TGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGGCCTGGGTGC 5242

Qy       2270 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCC 2329
Db       5243 CCGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCCTGGTGAGCGCCGGCATCC 5302

Qy       2330 GCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381
Db       5303 GCAAGGTGCTGTTCCTGGACGGCATCGACAAGGCCCAGGAGGAGCACGAGAA 5354


Search completed: December 30, 2005, 19:16:36
Job time : 11936.2 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 1293.51 Seconds
(without alignments)
12659.489 Million cell updates/sec

Title: US-09-610-313B-32
Perfect score: 2457
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2457	100.0	2457	6	ABL39961	Abl39961 Synthetic
2	2457	100.0	2457	12	ADM73766	Adm73766 HIV-1 pol
3	2443.4	99.4	2445	8	ACA03546	Aca03546 Synthetic
4	2443.4	99.4	2445	10	ADC13264	Adc13264 DNA of HI
5	2436.4	99.2	3930	10	ADC13230	Adc13230 DNA of HI
6	2436.2	99.2	2463	6	ABL39960	Abl39960 Synthetic
7	2436.2	99.2	2463	12	ADM73765	Adm73765 HIV-1 pol
8	2434.8	99.1	3930	10	ADC13231	Adc13231 DNA of HI
9	2434.8	99.1	3930	10	ADC13232	Adc13232 DNA of HI
10	2434.8	99.1	5184	8	ACA03591	Aca03591 Synthetic
11	2434.8	99.1	5184	10	ADC13279	Adc13279 DNA of HI
12	2428.6	98.8	2457	8	ACA03547	Aca03547 Synthetic
13	2428.6	98.8	2457	10	ADC13265	Adc13265 DNA of HI
14	2415.4	98.3	2469	6	ABL39959	Abl39959 Synthetic
15	2415.4	98.3	2469	12	ADM73764	Adm73764 HIV-1 pol
16	2404.4	97.9	3531	10	ADC13234	Adc13234 DNA of HI
17	2403.4	97.8	2457	8	ACA03548	Aca03548 Synthetic
18	2403.4	97.8	2457	10	ADC13266	Adc13266 DNA of HI
19	2402.8	97.8	3537	10	ADC13236	Adc13236 DNA of HI
20	2401.8	97.8	5145	8	ACA03521	Aca03521 Synthetic
21	2401.8	97.8	5145	10	ADC13233	Adc13233 DNA of HI
22	2391.8	97.3	3538	10	ADC13235	Adc13235 DNA of HI
23	2343.4	95.4	3607	8	ACA03551	Aca03551 Synthetic
24	2343.4	95.4	3607	10	ADC13269	Adc13269 DNA of HI
25	2325.2	94.6	3597	8	ACA03549	Aca03549 Synthetic
26	2325.2	94.6	3597	10	ADC13267	Adc13267 DNA of HI
27	2295.8	93.4	3624	8	ACA03550	Aca03550 Synthetic
28	2295.8	93.4	3624	10	ADC13268	Adc13268 DNA of HI
29	2136	86.9	2460	8	ACA03541	Aca03541 Synthetic
30	2136	86.9	2460	8	ACC78505	Acc78505 HIV p2Pol
31	2135	86.9	3564	8	ACC78488	Acc78488 HIV GagPo
32	2135	86.9	3564	8	ACC78489	Acc78489 HIV GagPo
33	2134.4	86.9	4716	8	ACA03522	Aca03522 Synthetic
34	2134.4	86.9	4716	10	ADC13238	Adc13238 DNA of HI
35	2130.6	86.7	3999	8	ACC78484	Acc78484 HIV GagCo
36	2129	86.7	3999	8	ACC78485	Acc78485 HIV GagCo
37	2129	86.7	3999	8	ACC78486	Acc78486 HIV GagCo
38	2129	86.7	5283	8	ACA03584	Aca03584 Synthetic
39	2129	86.7	5283	8	ACC78529	Acc78529 HIV TatRe
40	2128.8	86.6	4713	8	ACA03592	Aca03592 Synthetic
41	2128.8	86.6	4713	10	ADC13280	Adc13280 DNA of HI
42	2128.4	86.6	3462	10	ADC13237	Adc13237 DNA of HI
43	2115.2	86.1	2466	8	ACA03542	Aca03542 Synthetic
44	2115.2	86.1	2466	8	ACC78506	Acc78506 HIV p2Pol
45	2103	85.6	3735	8	ACA03545	Aca03545 Synthetic

ALIGNMENTS

RESULT 1
ABL39961
ID ABL39961 standard; DNA; 2457 BP.
XX
AC ABL39961;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975YMWM SEQ ID NO:32.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 10; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef

CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 100.0%; Score 2457; DB 6; Length 2457;
Best Local Similarity 100.0%; Pred. No. 1.1e-295;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
```

```
Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340

Qy       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400

Qy       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2457
```

RESULT 2
ADM73766
ID ADM73766 standard; DNA; 2457 BP.
XX
AC ADM73766;
XX
DT 03-JUN-2004 (first entry)
XX
DE HIV-1 polynucleotide #9.
XX
KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW HIV type C protein; immunostimulant.
XX
OS Human immunodeficiency virus 1.
XX
PN US2003223961-A1.
XX
PD 04-DEC-2003.
XX
PF 05-JUL-2001; 2001US-00899575.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (MEGE/) MEGEDE J Z.
PA (BARN/) BARNETT S W.
PA (ENGE/) ENGELBRECHT S.
PA (RENS/) RENSBURG E J V.
XX
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;
XX
DR WPI; 2004-060515/06.
XX
PT New expression cassette comprising a polynucleotide sequence encoding an
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT immunization, generating of packaging cell lines or in producing HIV Type
PT C proteins.
XX
PS Claim 1; SEQ ID NO 32; 160pp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV Type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
XX
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 100.0%; Score 2457; DB 12; Length 2457;
Best Local Similarity 100.0%; Pred. No. 1.1e-295;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACC 1380

Qy       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440

Qy       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1500

Qy       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAG 1560

Qy       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620

Qy       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAG 1680

Qy       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740

Qy       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATC 1800

Qy       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACC 1860

Qy       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCC 1920

Qy       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTC 1980

Qy       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040

Qy       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAG 2100

Qy       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAG 2160

Qy       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 TACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAG 2220

Qy       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 ATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAG 2280

Qy       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340

Qy       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTG 2400

Qy       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGGCTAGCACCGGTGAATTC 2457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGGCTAGCACCGGTGAATTC 2457
```

RESULT 3
ACA03546
ID ACA03546 standard; DNA; 2445 BP.
XX
AC ACA03546;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA ecoding immunogenic HIV peptide #29.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C

PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 34; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 99.4%; Score 2443.4; DB 8; Length 2445;
Best Local Similarity 100.0%; Pred. No. 5.3e-294;
Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
Db       1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy      67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
Db      61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy     127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
Db     121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy     187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
Db     181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy     247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 306
Db     241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGC 300

Qy     307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
Db     301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy     367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
Db     361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy     427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
Db     421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy     487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
Db     481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy     547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
Db     541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy     607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
Db     601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy     667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
Db     661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720

Qy     727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db     721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780

Qy     787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db     781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840

Qy     847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db     841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900

Qy     907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db     901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960

Qy     967 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db     961 GGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020

Qy    1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db    1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080

Qy    1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db    1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140

Qy    1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db    1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200

Qy    1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
Db    1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260

Qy    1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db    1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320

Qy    1327 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1386
Db    1321 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1380

Qy    1387 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1446
Db    1381 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1440

Qy    1447 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506
Db    1441 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500

Qy    1507 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1566
Db    1501 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1560

Qy    1567 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1626
Db    1561 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1620

Qy    1627 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1686
Db    1621 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1680

Qy    1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1746
Db    1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1740

Qy    1747 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1806
Db    1741 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1800

Qy    1807 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1866
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1860

Qy       1867 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1926
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1920

Qy       1927 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1986
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1980

Qy       1987 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2046
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2040

Qy       2047 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2106
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2100

Qy       2107 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2166
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2160

Qy       2167 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2226
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2220

Qy       2227 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2286
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2280

Qy       2287 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2346
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2340

Qy       2347 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2406
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2400

Qy       2407 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
                |||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445


RESULT 4
ADC13264
ID   ADC13264 standard; DNA; 2445 BP.
XX
AC   ADC13264;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YMWM_C SEQ ID NO 43.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 40; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

  Query Match           99.4%;  Score 2443.4;  DB 10;  Length 2445;
  Best Local Similarity  100.0%;  Pred. No. 5.3e-294;
  Matches 2444;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGC 60

Qy         67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120

Qy        127 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCAC 180

Qy        187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCC 240

Qy        247 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACCAGCCGC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACCAGCCGC 300

Qy        307 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACC 360

Qy        367 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGC 420

Qy        427 CAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 486
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        421 CAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGC 480

Qy        487 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGC 540

Qy        547 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATC 600

Qy        607 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTG 660

Qy        667 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 726
```

```
Db       661 AACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGC 720
Qy       727 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 786
Db       721 CCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC 780
Qy       787 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 846
Db       781 GAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACC 840
Qy       847 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db       841 CCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900
Qy       907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
Db       901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 960
Qy       967 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1026
Db       961 GGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTG 1020
Qy      1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1086
Db      1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAG 1080
Qy      1087 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db      1081 ACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy      1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db      1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy      1207 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1266
Db      1201 GTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCC 1260
Qy      1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1326
Db      1261 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAG 1320
Qy      1327 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1386
Db      1321 CACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAG 1380
Qy      1387 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1446
Db      1381 CCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGC 1440
Qy      1447 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506
Db      1441 AAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500
Qy      1507 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1566
Db      1501 CTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAG 1560
Qy      1567 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1626
Db      1561 CTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGC 1620
Qy      1627 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1686
Db      1621 AAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTAC 1680
Qy      1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1746
Db      1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACC 1740
Qy      1747 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1806
Db      1741 AACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATC 1800
Qy      1807 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1866
Db      1801 TGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1860
Qy      1867 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1926
Db      1861 ACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTG 1920
Qy      1927 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1986
Db      1921 GTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTG 1980
Qy      1987 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2046
Db      1981 GACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGC 2040
Qy      2047 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2106
Db      2041 CGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATC 2100
Qy      2107 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2166
Db      2101 CAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2160
Qy      2167 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2226
Db      2161 CTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATC 2220
Qy      2227 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2286
Db      2221 GAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATC 2280
Qy      2287 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2346
Db      2281 GGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTG 2340
Qy      2347 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2406
Db      2341 GACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGC 2400
Qy      2407 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db      2401 GGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2445
```

```
RESULT 5
ADC13230
ID   ADC13230 standard; DNA; 3930 BP.
XX
AC   ADC13230;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmut_C SEQ ID NO 9.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
```

PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 6; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramucosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 99.2%; Score 2436.4; DB 10; Length 3930;
Best Local Similarity 100.0%; Pred. No. 3.6e-293;
Matches 2437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226
```

```
Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCTGGTGAAGC 3406

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             |||||||||||||||||||||||||||||||||||||||
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 6
ABL39960
ID   ABL39960 standard; DNA; 2463 BP.
XX
AC   ABL39960;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Synthetic construct PR975YM SEQ ID NO:31.
XX
KW   Human immunodeficiency virus type C; antigenic HIV type C protein;
KW   immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW   immunostimulant; gene therapy; gene; ds.
XX
OS   Human immunodeficiency virus; type C.
OS   Synthetic.
XX
PN   WO200204493-A2.
XX
PD   17-JAN-2002.
XX
PF   05-JUL-2001; 2001WO-US021241.
XX
PR   05-JUL-2000; 2000US-00610313.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2002-154920/20.
XX
PT   New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT   applications including DNA immunization or generation of packaging cell
PT   lines, particularly in gene therapy.
XX
PS   Claim 1; Fig 9; 233pp; English.
XX
CC   The present invention describes expression cassettes comprising a
CC   polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC   type C polypeptides. The expression cassettes comprise any of the HIV
CC   type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC   (I). (I) have immunostimulant activity and can be used in gene therapy.
CC   The HIV type C polynucleotides are useful in applications including DNA
CC   immunisation, generation of packaging cell lines, and production of HIV
CC   Type C proteins. The polynucleotides are particularly useful in gene
CC   therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC   ABB06204 to ABB06215 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

  Query Match             99.2%;  Score 2436.2;  DB 6;  Length 2463;
  Best Local Similarity   99.6%;  Pred. No. 4.1e-293;
  Matches 2454;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;
```

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
```

```
Qy       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db       601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db       661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db       721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db       781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db       841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db       901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db       961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db      1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db      1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db      1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy      1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db      1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db      1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
Db      1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy      1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
Db      1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy      1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
Db      1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy      1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
Db      1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy      1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
Db      1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy      1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
Db      1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy      1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
Db      1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy      1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
Db      1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy      1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
Db      1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy      1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
Db      1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy      1915 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
Db      1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy      1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
Db      1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy      2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
Db      2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy      2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
Db      2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy      2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
Db      2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy      2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
Db      2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy      2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
Db      2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy      2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
Db      2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy      2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454
Db      2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy      2455 TTC 2457
Db      2461 TTC 2463
```

```
RESULT 7
ADM73765
ID   ADM73765 standard; DNA; 2463 BP.
XX
AC   ADM73765;
XX
DT   03-JUN-2004  (first entry)
XX
DE   HIV-1 polynucleotide #8.
XX
KW   HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW   HIV type C protein; immunostimulant.
XX
OS   Human immunodeficiency virus 1.
XX
PN   US2003223961-A1.
XX
```

PD 04-DEC-2003.
XX
PF 05-JUL-2001; 2001US-00899575.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (MEGE/) MEGEDE J Z.
PA (BARN/) BARNETT S W.
PA (ENGE/) ENGELBRECHT S.
PA (RENS/) RENSBURG E J V.
XX
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;
XX
DR WPI; 2004-060515/06.
XX
PT New expression cassette comprising a polynucleotide sequence encoding an
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT immunization, generating of packaging cell lines or in producing HIV Type
PT C proteins.
XX
PS Claim 1; SEQ ID NO 31; 160pp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV Type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
XX
SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 99.2%; Score 2436.2; DB 12; Length 2463;
Best Local Similarity 99.6%; Pred. No. 4.1e-293;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140

Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
              |||||||||||||||||||||||||||||||||||||    ||      ||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
```

```
Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460

Qy       2455 TTC 2457
             |||
Db       2461 TTC 2463


RESULT 8
ADC13231
ID   ADC13231 standard; DNA; 3930 BP.
XX
AC   ADC13231;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutAtt_C SEQ ID NO 10.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 7; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;

  Query Match           99.1%;  Score 2434.8;  DB 10;  Length 3930;
  Best Local Similarity   99.9%;  Pred. No. 5.7e-293;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846
```

```
Qy       374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db      1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy       434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db      1907 AGGAGGCCCTGCTGGACTCCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy       494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db      1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy       554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db      2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy       614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db      2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy       674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db      2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy       734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db      2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy       794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db      2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy       854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db      2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy       914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db      2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy       974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db      2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy      1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db      2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy      1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db      2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy      1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db      2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy      1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db      2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy      1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db      2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy      1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
Db      2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy      1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
Db      2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy      1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
Db      2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy      1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
Db      2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy      1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
Db      3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy      1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
Db      3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy      1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
Db      3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy      1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
Db      3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy      1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
Db      3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy      1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
Db      3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy      1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
Db      3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy      1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
Db      3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy      2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
Db      3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy      2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
Db      3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy      2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
Db      3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy      2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
Db      3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy      2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
Db      3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy      2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
Db      3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy      2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db      3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

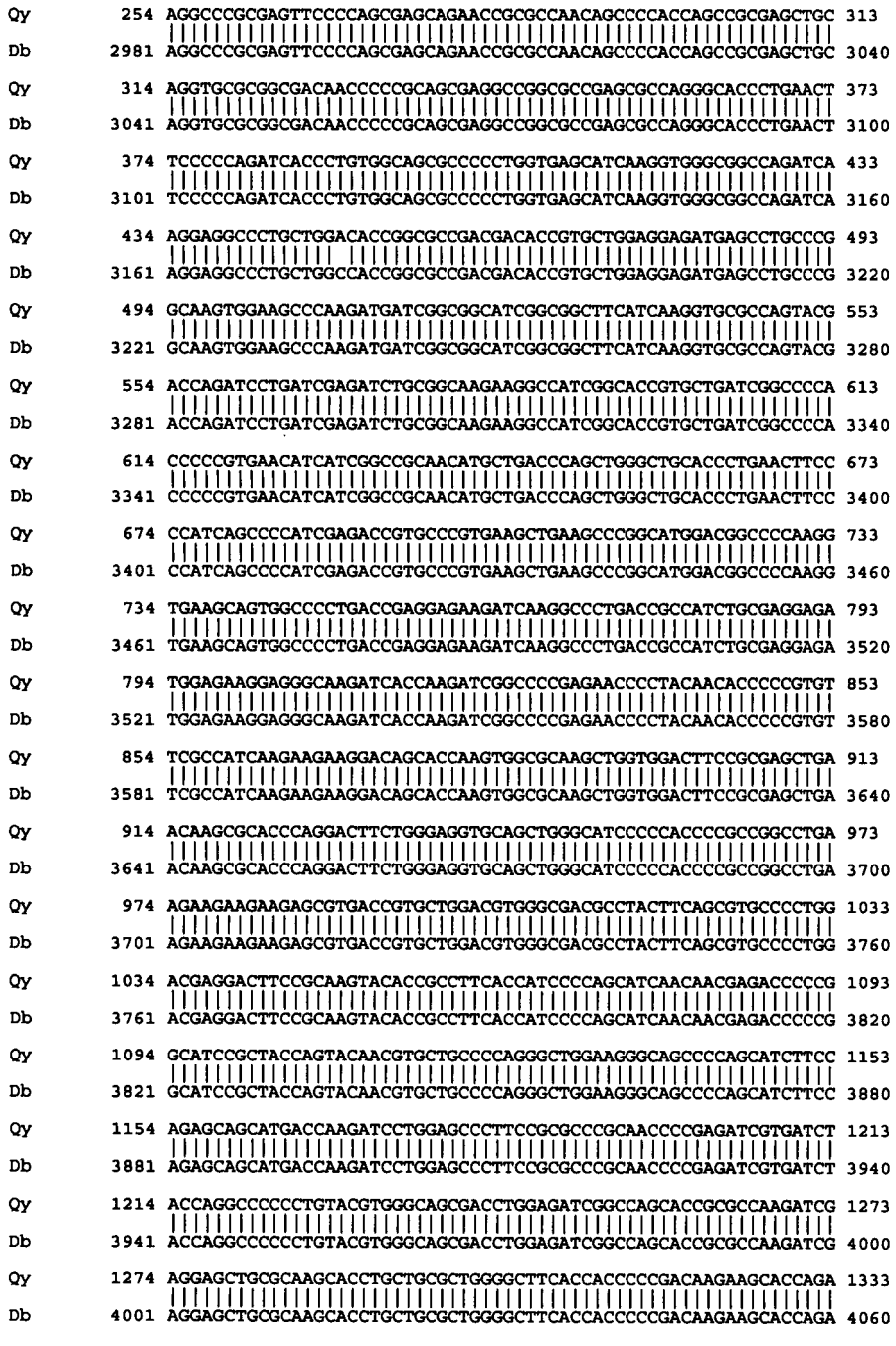
RESULT 9
ADC13232
ID ADC13232 standard; DNA; 3930 BP.
XX
AC ADC13232;

```
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct GagComplPolmutIna_C SEQ ID NO 11.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 8; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;

  Query Match            99.1%;  Score 2434.8;  DB 10;  Length 3930;
  Best Local Similarity  99.9%;  Pred. No. 5.7e-293;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 1546

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 1606

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1607 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 1666

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 1726

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 1786

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1847 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 1906

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       1907 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 1966

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 2026

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 2086

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 2146

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 2206

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2267 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 2326

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2327 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 2446

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 2506

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 2566

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 2746

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 2806

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2807 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 2866

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1787 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 1846
```

```
Db       2867 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 2926

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2927 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 2986

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2987 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 3046

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3047 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 3106

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3107 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 3166

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 3226

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 3286

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 3406

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 3466

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 3526

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 3586

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 3646

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 3706

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3707 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 3766

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 3886

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             |||||||||||||||||||||||||||||||||||||
Db       3887 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

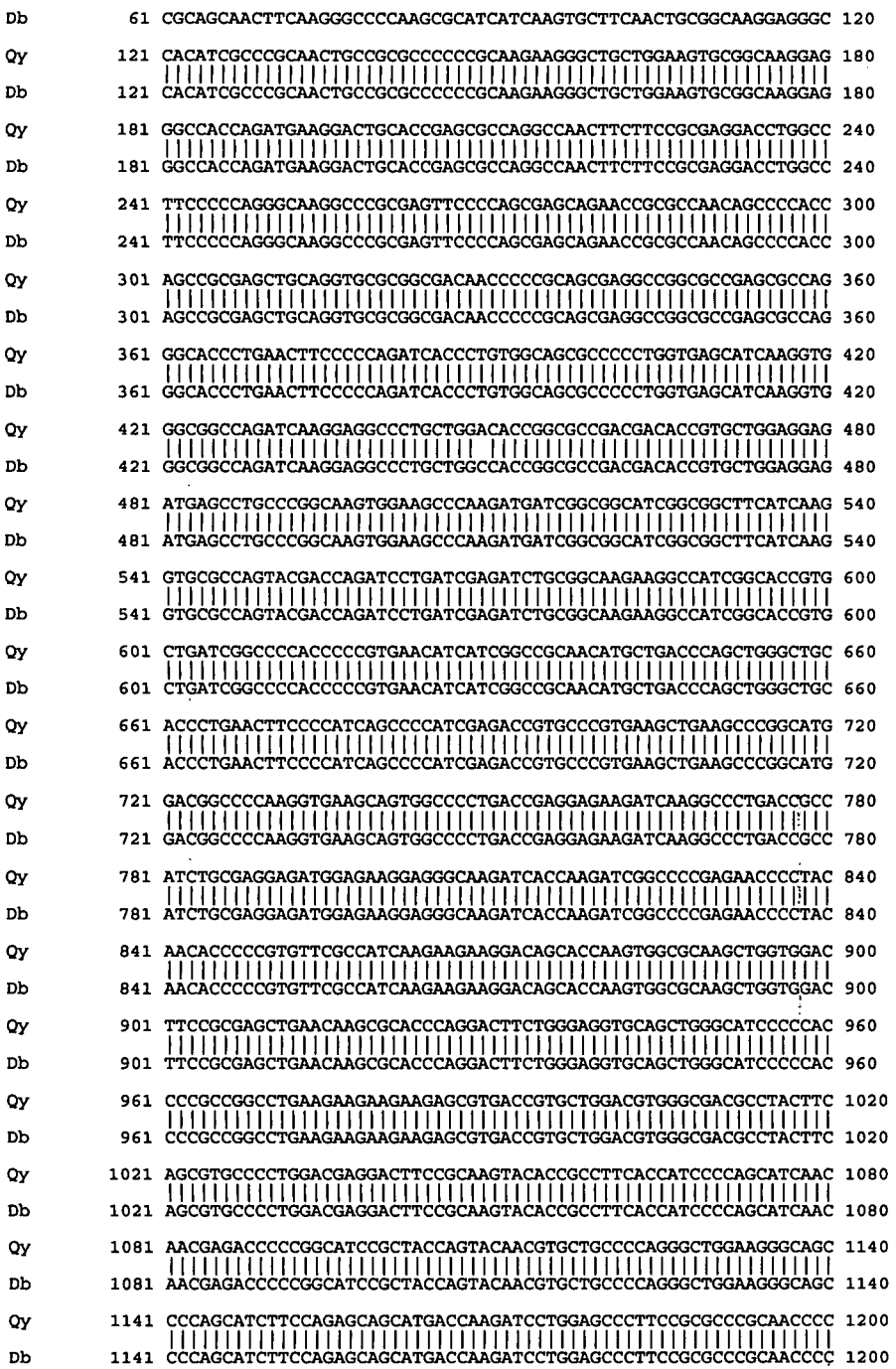
RESULT 10
ACA03591

```
ID   ACA03591 standard; DNA; 5184 BP.
XX
AC   ACA03591;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #74.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 79; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

  Query Match             99.1%;  Score 2434.8;  DB 8;  Length 5184;
  Best Local Similarity   99.9%;  Pred. No. 5.5e-293;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACTGCCGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980
```


Qy 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
Db 2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040
Qy 314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
Db 3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100
Qy 374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
Db 3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
Db 3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220
Qy 494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
Db 3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280
Qy 554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
Db 3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340
Qy 614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
Db 3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400
Qy 674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
Db 3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460
Qy 734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520
Qy 794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
Db 3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580
Qy 854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
Db 3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700
Qy 974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
Db 3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760
Qy 1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
Db 3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820
Qy 1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
Db 3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940
Qy 1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
Db 3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000
Qy 1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
Db 4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060
Qy 1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
Db 4061 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 4120
Qy 1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
Db 4121 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 4180
Qy 1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
Db 4181 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 4240
Qy 1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
Db 4241 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 4300
Qy 1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
Db 4301 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 4360
Qy 1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
Db 4361 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 4420
Qy 1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
Db 4421 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 4480
Qy 1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
Db 4481 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 4540
Qy 1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
Db 4541 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 4600
Qy 1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
Db 4601 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 4660
Qy 1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
Db 4661 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 4720
Qy 1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
Db 4721 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 4780
Qy 2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
Db 4781 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 4840
Qy 2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
Db 4841 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 4900
Qy 2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
Db 4901 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 4960
Qy 2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
Db 4961 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 5020
Qy 2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
Db 5021 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 5080
Qy 2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
Db 5081 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 5140
Qy 2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451

```
                ||||||||||||||||||||||||||||||||||||||
Db       5141 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178


RESULT 11
ADC13279
ID   ADC13279 standard; DNA; 5184 BP.
XX
AC   ADC13279;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct TatRevNefgagCpo1Ina_C SEQ ID NO 58.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 55; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

  Query Match             99.1%;  Score 2434.8;  DB 10;  Length 5184;
  Best Local Similarity   99.9%;  Pred. No. 5.5e-293;
  Matches 2436;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 73
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAGCGCAGCAACTTCA 2800

Qy         74 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860

Qy        134 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 193
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGA 2920

Qy        194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 253
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCA 2980

Qy        254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 313
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGC 3040

Qy        314 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 373
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 AGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACT 3100

Qy        374 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 433
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 TCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCA 3160

Qy        434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 493
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       3161 AGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCG 3220

Qy        494 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 553
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 GCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACG 3280

Qy        554 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 613
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 ACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCA 3340

Qy        614 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 673
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 CCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCC 3400

Qy        674 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 733
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 CCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG 3460

Qy        734 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 TGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520

Qy        794 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 853
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 TGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGT 3580

Qy        854 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 TCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy        914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 973
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGA 3700

Qy        974 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 1033
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 AGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGG 3760

Qy       1034 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 1093
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 ACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGAGACCCCCG 3820

Qy       1094 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 GCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 3880

Qy       1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3881 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 3940

Qy       1214 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 1273
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3941 ACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCG 4000

Qy       1274 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 1333
```

```
Db       4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGA 4060

Qy       1334 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 1393
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4061 AGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCG 4120

Qy       1394 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4121 AGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGA 4180

Qy       1454 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4181 ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 4240

Qy       1514 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4241 GCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCG 4300

Qy       1574 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 AGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAGGACC 4360

Qy       1634 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 TGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGC 4420

Qy       1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4421 CCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACG 4480

Qy       1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4481 TGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA 4540

Qy       1814 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4541 AGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 4600

Qy       1874 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4601 ACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGC 4660

Qy       1934 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 1993
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4661 TGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCG 4720

Qy       1994 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 2053
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4721 CCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGA 4780

Qy       2054 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 2113
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4781 AGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAGCTGCAGGCCATCCAGCTGG 4840

Qy       2114 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 2173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4841 CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA 4900

Qy       2174 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 2233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4901 TCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGC 4960

Qy       2234 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 2293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4961 TGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCA 5020

Qy       2294 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5021 ACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 5080

Qy       2354 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 2413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       5081 TCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCC 5140

Qy       2414 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             |||||||||||||||||||||||||||||||||||||
Db       5141 CTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 5178


RESULT 12
ACA03547
ID   ACA03547 standard; DNA; 2457 BP.
XX
AC   ACA03547;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Synthetic DNA ecoding immunogenic HIV peptide #30.
XX
KW   Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW   gene therapy; packaging cell line; humoral immune response;
KW   cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS   Synthetic.
XX
PN   WO2003004657-A1.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021421.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349728P.
PR   16-JAN-2002; 2002US-0349793P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y;
XX
DR   WPI; 2003-221602/21.
XX
PT   New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT   polypeptides, useful as immunogenic compositions or vaccines for
PT   generating humoral or cellular immune responses against HIV in a subject,
PT   especially humans.
XX
PS   Example 1; Fig 35; 262pp; English.
XX
CC   The invention describes a synthetic polynucleotide encoding 2 or more
CC   immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC   derived from different HIV subtypes. The polynucleotide is useful for
CC   immunisation, generation of packaging cell lines, or production of HIV
CC   polypeptides. The polynucleotide and its encoded proteins are useful as
CC   immunogenic compositions or vaccines for generating humoral or cellular
CC   immune responses against HIV in a subject, or for inducing neutralising
CC   antibodies against HIV. The gene delivery vector comprising the
CC   polynucleotide is also useful for DNA immunisation of, or for generating
CC   an immune response (e.g. a humoral or cellular immune response) in, a
CC   subject such as a mammal, particularly a human. This sequence encodes a
CC   human immunodeficiency virus immunogenic peptide
XX
SQ   Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match            98.8%;  Score 2428.6;  DB 8;  Length 2457;
  Best Local Similarity   99.6%;  Pred. No. 3.6e-292;
  Matches 2447;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```


Db 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Qy 121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db 121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Qy 181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db 181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Qy 241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300
Db 241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Qy 301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db 301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Qy 361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db 361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Qy 421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db 421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Qy 481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db 481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Qy 541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db 541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Qy 601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Qy 661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db 661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Qy 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Qy 781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db 781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Qy 841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db 841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy 961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db 961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy 1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Db 1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
Qy 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Db 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
Qy 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
Db 1321 AAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy 1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy 1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
Db 1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy 1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
Db 1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560
Qy 1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
Db 1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620
Qy 1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
Db 1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680
Qy 1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
Db 1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy 1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
Db 1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800
Qy 1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
Db 1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860
Qy 1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
Db 1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920
Qy 1915 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
Db 1921 ACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980
Qy 1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
Db 1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040
Qy 2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
Db 2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100
Qy 2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
Db 2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160
Qy 2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
Db 2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220
Qy 2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
Db 2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

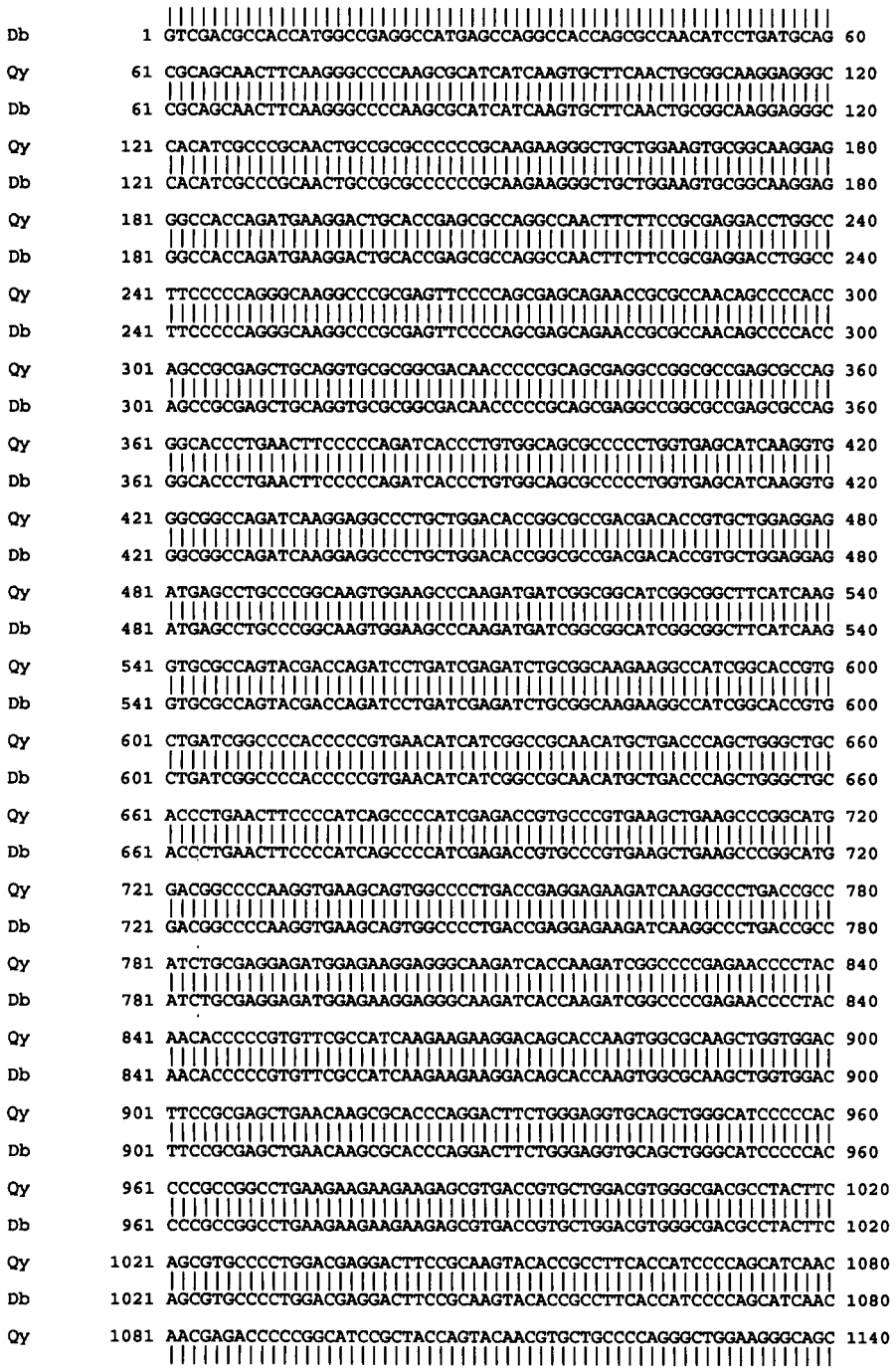
```
Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457


RESULT 13
ADC13265
ID   ADC13265 standard; DNA; 2457 BP.
XX
AC   ADC13265;
XX
DT   18-DEC-2003  (first entry)
XX
DE   DNA of HIV construct p2Pol-opt-YM_C SEQ ID NO 44.
XX
KW   expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW   Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS   Human immunodeficiency virus.
XX
PN   WO2003004620-A2.
XX
PD   16-JAN-2003.
XX
PF   05-JUL-2002; 2002WO-US021420.
XX
PR   05-JUL-2001; 2001US-0303192P.
PR   31-AUG-2001; 2001US-0316860P.
PR   16-JAN-2002; 2002US-0349871P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (UYST-) UNIV STELLENBOSCH.
XX
PI   Zur Megede J,  Barnett SW,  Lian Y,  Engelbrecht S,  Van Rensburg EJ;
XX
DR   WPI; 2003-221593/21.
XX
PT   New expression cassette comprising a polynucleotide sequence encoding a
PT   polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT   Prot, or Rev polypeptide, useful for immunization, or generating
PT   packaging cell lines.
XX
PS   Disclosure; Fig 41; 301pp; English.
XX
CC   The invention relates to a novel expression cassette comprising a
CC   polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC   Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC   expression cassette can be used to treat HIV type C by gene therapy or
CC   used in the development of a vaccine. The gene delivery vector is
CC   administered intramuscularly, intramucosally, intranasally,
CC   subcutaneously, intradermally, transdermally, intravaginally,
CC   intrarectally, orally or intravenously. The expression cassette is useful
CC   for immunisation, generating packaging cell lines and producing HIV
CC   polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC   Type C related sequence of the invention.
XX
SQ   Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

  Query Match           98.8%;  Score 2428.6;  DB 10;  Length 2457;
  Best Local Similarity  99.6%;  Pred. No. 3.6e-292;
  Matches 2447;  Conservative    0;  Mismatches    4;  Indels    6;  Gaps    1;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGCCACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020

Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080

Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
```

```
Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200

Qy       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCAC 1260

Qy       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGAC 1320

Qy       1321 AAGAAGCACCAGAAGGAGCCCCCCCTTCCTGCCCAT------CGAGCTGCACCCCGACAAG 1374
             |||||||||||||||||||||||||||||||    ||      ||||||||||||||||||||
Db       1321 AAGAAGCACCAGAAGGAGCCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380

Qy       1375 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAG 1440

Qy       1435 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500

Qy       1495 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1554
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAG 1560

Qy       1555 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1614
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTAC 1620

Qy       1615 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1680

Qy       1675 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740

Qy       1735 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1794
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAG 1800

Qy       1795 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1854
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAGACCTGG 1860

Qy       1855 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1914
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAAC 1920

Qy       1915 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1974
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAG 1980

Qy       1975 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 ACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTG 2040

Qy       2035 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2094
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAGACCGAG 2100

Qy       2095 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGAC 2160

Qy       2155 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTG 2220

Qy       2215 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCC 2280

Qy       2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340

Qy       2335 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGACGACCTG 2400

Qy       2395 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2451
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 14
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 8; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (I) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 98.3%; Score 2415.4; DB 6; Length 2469;
Best Local Similarity 99.3%; Pred. No. 1.5e-290;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
```


Db 1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
Qy 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Db 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
Qy 121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Db 121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
Qy 181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db 181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
Qy 241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Db 241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
Qy 301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Db 301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
Qy 361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db 361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
Qy 421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Db 421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
Qy 481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db 481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Qy 541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Db 541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
Qy 601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Qy 661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Db 661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
Qy 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Qy 781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Db 781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
Qy 841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db 841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
Qy 961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Db 961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy 1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
Db 1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Qy 1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
Db 1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Qy 1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368
Db 1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Qy 1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
Db 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Qy 1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
Db 1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Qy 1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
Db 1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Qy 1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
Db 1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Qy 1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
Db 1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Qy 1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
Db 1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Qy 1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
Db 1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Qy 1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
Db 1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Qy 1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
Db 1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Qy 1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
Db 1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Qy 1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
Db 1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Qy 2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
Db 2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Qy 2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
Db 2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Qy 2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
Db 2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220

```
Qy       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280

Qy       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340

Qy       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400

Qy       2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460

Qy       2449 GGTGAATTC 2457
             |||||||||
Db       2461 GGTGAATTC 2469
```

RESULT 15
ADM73764
ID ADM73764 standard; DNA; 2469 BP.
XX
AC ADM73764;
XX
DT 03-JUN-2004 (first entry)
XX
DE HIV-1 polynucleotide #7.
XX
KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW HIV type C protein; immunostimulant.
XX
OS Human immunodeficiency virus 1.
XX
PN US2003223961-A1.
XX
PD 04-DEC-2003.
XX
PF 05-JUL-2001; 2001US-00899575.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (MEGE/) MEGEDE J Z.
PA (BARN/) BARNETT S W.
PA (ENGE/) ENGELBRECHT S.
PA (RENS/) RENSBURG E J V.
XX
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;
XX
DR WPI; 2004-060515/06.
XX
PT New expression cassette comprising a polynucleotide sequence encoding an
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT immunization, generating of packaging cell lines or in producing HIV Type
PT C proteins.
XX
PS Claim 1; SEQ ID NO 30; 160pp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV Type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
XX
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 98.3%; Score 2415.4; DB 12; Length 2469;
Best Local Similarity 99.3%; Pred. No. 1.5e-290;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720

Qy        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840

Qy        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAC 960

Qy        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
```

```
Db        961 CCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTC 1020
Qy       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Db       1021 AGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAAC 1080
Qy       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Db       1081 AACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGC 1140
Qy       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Db       1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy       1201 GAGATCGTGATCTACCA------GGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGC 1254
Db       1201 GAGATCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGACCTGGAGATCGGC 1260
Qy       1255 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1314
Db       1261 CAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACC 1320
Qy       1315 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCAT------CGAGCTGCACCCC 1368
Db       1321 CCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTACGAGCTGCACCCC 1380
Qy       1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1428
Db       1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440
Qy       1429 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1488
Db       1441 ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Qy       1489 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1548
Db       1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACC 1560
Qy       1549 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1608
Db       1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGC 1620
Qy       1609 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1668
Db       1621 GTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1680
Qy       1669 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
Db       1681 TGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Qy       1729 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1788
Db       1741 ATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCC 1800
Qy       1789 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1848
Db       1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCAGAAGGAG 1860
Qy       1849 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1908
Db       1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Qy       1909 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1968
Db       1921 GTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATCATCGGC 1980
Qy       1969 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2028
Db       1981 GCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGGC 2040
Qy       2029 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2088
Db       2041 TACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAACCAGAAG 2100
Qy       2089 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2148
Db       2101 ACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTG 2160
Qy       2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2208
Db       2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGAGAGCGAG 2220
Qy       2209 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2268
Db       2221 CTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTG 2280
Qy       2269 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328
Db       2281 CCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Qy       2329 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2388
Db       2341 CGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTACATGGAC 2400
Qy       2389 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2448
Db       2401 GACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACC 2460
Qy       2449 GGTGAATTC 2457
Db       2461 GGTGAATTC 2469
```

Search completed: December 30, 2005, 08:57:08
Job time : 1300.51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:41 ; Search time 8676.48 Seconds
(without alignments)
13249.138 Million cell updates/sec

Title: US-09-610-313B-32
Perfect score: 2457
Sequence: 1 gtcgacgccaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	136.4	5.6	330	10	CL293849	CL293849 02S0349-0
	2	100	4.1	2886	10	CL967755	CL967755 OsIFCC015
	3	90.2	3.7	1781	8	CX099623	CX099623 RECm0260
	4	90.2	3.7	2598	4	AY103647	AY103647 Zea mays
	5	88.8	3.6	3069	10	CL973991	CL973991 OsIFCC025
	6	86.6	3.5	1398	10	CL961989	CL961989 OsIFCC006
	7	86	3.5	743	10	CZ247380	CZ247380 AIAA-aaf3
	8	85.6	3.5	951	3	BM321451	BM321451 rockefell
	9	84.2	3.4	1941	10	CL971508	CL971508 OsIFCC021
c	10	80.8	3.3	1060	10	CW922203	CW922203 EDCAR29TR
c	11	80.6	3.3	869	7	CK159167	CK159167 FGAS04056
	12	80.6	3.3	892	10	CZ216254	CZ216254 AIAA-aaf2
	13	80.2	3.3	1132	3	BM320864	BM320864 rockefell
	14	79.6	3.2	1165	3	BM320900	BM320900 rockefell
	15	79.6	3.2	1680	10	CL982770	CL982770 OsIFSC049
	16	79	3.2	867	3	BM321430	BM321430 rockefell
	17	78.4	3.2	1725	10	CL978463	CL978463 OsIFCC031
	18	78.4	3.2	2031	10	CL974989	CL974989 OsIFCC042
	19	78.2	3.2	2697	10	CL952258	CL952258 OsJRUA000
	20	78.2	3.2	2853	10	CL974397	CL974397 OsIFCC025
	21	78	3.2	1485	10	CL970981	CL970981 OsIFCC020
	22	77.4	3.2	1509	10	CL959255	CL959255 OsIFCC002
	23	77.4	3.2	2559	10	CL982027	CL982027 OsIFSC046
	24	77	3.1	11691	10	CL962901	CL962901 OsIFCC008
	25	76.8	3.1	2682	10	CL969033	CL969033 OsIFCC017
	26	76.6	3.1	1550	3	BM321022	BM321022 rockefell
	27	76.2	3.1	1401	10	CL962721	CL962721 OsIFCC038
	28	76.2	3.1	3249	10	CL945510	CL945510 OsIFSB004
	29	75	3.1	545	3	BI724851	BI724851 1031075E0
	30	75	3.1	862	3	BM321023	BM321023 rockefell
	31	74.4	3.0	914	9	BZ568300	BZ568300 pacs2-164
	32	74.2	3.0	2313	10	CL982362	CL982362 OsIFSC047
	33	74	3.0	1290	10	CL972679	CL972679 OsIFCC023
	34	74	3.0	1386	11	DQ045165	DQ045165 Homo sapi
	35	74	3.0	2072	4	CR603312	CR603312 full-leng
	36	73.8	3.0	2433	10	AY401196	AY401196 Homo sapi
	37	73.2	3.0	853	3	BM321393	BM321393 rockefell
c	38	72.8	3.0	566	3	BM587428	BM587428 170006873
	39	72.6	3.0	718	4	CNS0GHPN	CR718981 Tetraodon
	40	72.6	3.0	788	6	CB643171	CB643171 OSJNEb03L
	41	72.4	2.9	753	9	CC675888	CC675888 OGWCO51TH
	42	72.4	2.9	2523	10	CL974879	CL974879 OsIFCC026
	43	72.2	2.9	602	7	CV057146	CV057146 BNEL24e3
	44	72.2	2.9	640	2	BE601575	BE601575 HVSMEh009
	45	72.2	2.9	688	6	CB648640	CB648640 OSJNEb12C

ALIGNMENTS

RESULT 1
CL293849/c
LOCUS CL293849 330 bp DNA linear GSS 12-FEB-2004
DEFINITION 02S0349-08A1-C03 UniformMu MuTAIL Library Zea mays genomic clone 02S0349-08A1-C03, genomic survey sequence.
ACCESSION CL293849
VERSION CL293849.1 GI:42541978
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 330)
AUTHORS Latshaw,S., Tan,B-.C., Settles,A.M. and McCarty,D.R.
TITLE Sequence tagged transposon insertions from the UniformMu maize population
JOURNAL Unpublished (2003)
COMMENT Contact: Donald R. McCarty
Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drm@ufl.edu
Sequence flanking probable Mu insertion site in UniformMu line: 02S0349-08, Primer set: A
Class: transposon insertion site.
FEATURES Location/Qualifiers
source 1. .330
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="02S0349-08A1-C03"
/clone_lib="UniformMu MuTAIL Library"
/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo assymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sepharose 400 spin columns and cloned into the TOPO PCR4 vector."
ORIGIN

```
  Query Match             5.6%;  Score 136.4;  DB 10;  Length 330;
  Best Local Similarity   66.4%;  Pred. No. 3.1e-14;
  Matches  211;  Conservative    0;  Mismatches  106;  Indels    1;  Gaps    1;

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              || || ||| |  | || || ||  | || || || || || || || ||  | || ||
Db        329 GGGGGGCAGCTGGAAGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAA 270

Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
              ||||   || | || |  ||||| || || ||||| || || || || || || |||||
Db        269 ATGACTTTGACAGGAAGATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAA 210

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTG 600
              ||  | ||||| || ||| | |  || || ||||| ||  | || || || || || ||
Db        209 GTAAGACAGTATGATCAGGTACCCATAGAAATCTGTGGGCATAAAGCTATAGGTACGGTA 150

Qy        601 CTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC 660
              |  | || || || || || ||||| || ||  | ||  || |||| ||| | || |||
Db        149 TTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGG-TGC 91

Qy        661 ACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
              ||| | || || ||||| || || || || || || || || ||  | ||||| || |||
Db         90 ACCTTAAATTTTCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATG 31

Qy        721 GACGGCCCCAAGGTGAAG 738
              || ||||| || || | |
Db         30 GATGGCCCAAAAGTAAGG 13


RESULT 2
CL967755
LOCUS       CL967755               2886 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC015718 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL967755
VERSION     CL967755.1  GI:52390149
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 2886)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .2886
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             4.1%;  Score 100;  DB 10;  Length 2886;
  Best Local Similarity   42.8%;  Pred. No. 1.2e-07;
  Matches  700;  Conservative    0;  Mismatches  915;  Indels   21;  Gaps    3;

Qy        189 GATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCA 248
              | | | |||   | |||   |  ||||| | || ||  | | | |   | |  || |

Db        219 GGTCATGGAGGCCGCCGCCATCATGGCCATCGTCCTCGCCAACGGCGGCGGCAGGCCGCC 278

Qy        249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGA 308
              | | |     |    ||  || ||        || || || | |||||| | |
Db        279 CGACTGGCAGGACTTCGTCGGCATCGTCGTCCTCCTCGTCATCAACTCCACCATCAGTTT 338

Qy        309 GCTGCAGGTGCGC---GGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCAC 365
              |  ||| |  |    | ||  |||||  |  |  |  |  ||||  |  |  ||  .
Db        339 CATCGAGGAGAACAATGCCGGGAACGCCGCCGCGGCGCTCATGGCCGGCCTCGCGCCCAA 398

Qy        366 CCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGG 425
              ||  | | |  |  |      ||||||  |    |  | ||||  ||  |||
Db        399 GACCAAGGTGCTCAGGGACGGGAAATGGCAGGAGCAGGACGCGTCCATCCTCGTTCCCGG 458

Qy        426 CCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAG 485
              | | ||||      | ||| | |  | |  | || | || || |  |  | | |
Db        459 CGACATCATCAGCATCAAGCTCGGCGACATCATCCCCGCCGACGCGCGGCTGCTCGAGGG 518

Qy        486 CCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCG 545
              |   ||  ||||    | |  |  | | ||  ||||          |  ||||   ||
Db        519 CGACCCGCTCAAGGTCGACCAGGCGGCGCTCACCGGCGAGTCGATGCCCGTCAACAAGCA 578

Qy        546 CCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGAT 605
              |       |  || |  || |  | || |||| ||| |    |  |  | |
Db        579 CGCCGGGCAGGGCGTCTTCTCCGGGTCCACCGTCAAGCAGGGCGAGATCGAGGCCGTCGT 638

Qy        606 CGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCT 665
              |  | |||||  |||| ||| | ||  | ||||  |  |  || | ||| | |||||
Db        639 CATCGCCACCGGCGTGCACACCTTCTTCGGCAAGGCGGCGCACCTGGTGGACAGCACCAA 698

Qy        666 GAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGG 725
              ||| ||  |  ||  ||   ||| |  |  |  |  ||  |               | ||| |
Db        699 CAACATCGGCCACTTCCAGCTCGTGCTCACGGCCATCGGCAACTTCTGCATCATCTCCAT 758

Qy        726 CCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTG 785
              |  |  | | |        | | ||  | |  |  |  |||  | |
Db        759 CGGCGTCGGCATGATAATCGAGATCATCGTTATGTACCCGATCCAGCACCGCGCGTACCG 818

Qy        786 CGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACAC 845
              ||| |  || || ||  |  |  |  ||  |||| |  |  | ||  |||
Db        819 CGACGGCATCGACAACCTGCTCGTCCTGCTCATCGGCGGCATCCCCATCGCCATGCCCAC 878

Qy        846 CCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCG 905
              |  ||| || || || |  |  |  |  |  |  |  |  |  ||| ||  | | ||
Db        879 CGTGCTGTCCGTCACCATGGCCATCGGCTCCCACCGGCTGTCGCAGCAGGGCGCCATCAC 938

Qy        906 CGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGC 965
              | ||| ||  |   || | ||||  |  |  || |  ||  |  |  | ||  |
Db        939 CAAGCGGATGACCGCCATCGAGGAGATGGCCGGCATGGACGTGCTCTGCAGCGACAAGAC 998

Qy        966 CGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTACTTCAGCGT 1025
              |||  |  || |  || |  | |||||  |  |  ||  | |  |
Db        999 CGGGACGCTGACGCTCAACAAGCTCACCGTCGACAAGACGCTGATCGAGGTGTACGGGCG 1058

Qy       1026 GCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATCAACAACGA 1085
              |  |||||| ||||||          |||  |      || || || | ||| |
Db       1059 GGGGCTGGACAAGGACTCGGTGCTCCTGTACGCGGCGAGGGCGTCCCGCGTCGAGAACCA 1118

Qy       1086 GACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAG 1145
              | | ||   |  | |       | |||| || |  |  | ||| | || || |
Db       1119 GGACGCCATTGACACATGCATCGTGGGCATGCTCGCCGACCCCAAGGAGGCCCGCGCCGG 1178

Qy       1146 CATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGAT 1205
              ||||    ||  |  | |  |           || |||  | ||  | |      |
Db       1179 CATCAAGGAGGTCCACTTCCTC-----------CCGTTCAACCCCGTCGAGAAGCGCAC 1226

Qy       1206 CGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCAGCACCGCGC 1265
              |  |||  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  | |
Db       1227 GGCCATCACCTACATCGACGGCAATGGCGAATGGCACAGGATCAGCAAGGGCGCGCCGGA 1286

Qy       1266 CAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCCCGACAAGAA 1325
              |||||   |||||| |||||          |  |  |  |  |  ||  ||
Db       1287 GCAGATCATCGAGCTGTGCAAGATGAGCAAGGACGCCGAGAAGAAGGTGCACACGCTGAT 1346
```

```
Qy       1326 GCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTGGACCGTGCA 1385
Db       1347 CGACCAGTACGCGGACCGCGGCCTCCGGTCGCTGGGCGTGTCGTACCAGAAGGTGCCGGA 1406

Qy       1386 GCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGG 1445
Db       1407 GAAGAGCAAGGAGAGCGAGGGCGAGCCATGGCAGTTCGTCGGCCTCCTGCCGCTGTTCGA 1466

Qy       1446 CAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCT 1505
Db       1467 CCCGCCGAGGCACGACAGCGCGGAGACCATCCGCCGCGCGCTGCACCTCGGCGTGAACGT 1526

Qy       1506 GCTGCGCGGCGCCAAGGCCCTGACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGA 1565
Db       1527 GAAGATGATCACCGGCGACCAGCTCGCCATCG------GGAAGGAGACGGCGCGGCGGCT 1580

Qy       1566 GCTGGCCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAG 1625
Db       1581 CGGGATGGGCACCAACATGTACCCGTCGACGACGCTGCTGGGCGACAAGAGCAGCGAGAT 1640

Qy       1626 CAAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTA 1685
Db       1641 GAGCGGCCTCCCAATCGACGAGCTGATCGAGAAGGCCGACGGCTTCGCCGGCGTGTTCCC 1700

Qy       1686 CCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACAC 1745
Db       1701 GGAGCACAAGTACGAGATCGTGAAGCGCCTCCAGGACCGGAAGCACATCTGCGGCATGAC 1760

Qy       1746 CAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGAT 1805
Db       1761 CGGCGACGGCGTGAACGACGCGCCGGCGCTGAAGAAGGCCGACATCGGCATCGCCGTCGA 1820

Qy       1806 CTGGGGCAAGACCCCC 1821
Db       1821 CGACGCCACGGACGCC 1836
```

RESULT 3
CX099623
LOCUS CX099623 1781 bp mRNA linear EST 03-JUN-2005
DEFINITION RECm0260 A normalized whole-life-cycle cDNA library of rice Oryza sativa (indica cultivar-group) cDNA clone EI057O14, THREEB09, EI057O04, EI062D05, BI027I14, EI070G16, EI08 5', mRNA sequence.
ACCESSION CX099623
VERSION CX099623.1 GI:66912775
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 1781)
AUTHORS Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D., Han,B., Zhang,Q. and Wang,S.
TITLE Features of the expressed sequences revealed by a large-scale analysis of ESTs from a normalized cDNA library of the elite indica rice cultivar Minghui 63
JOURNAL Plant J. 42 (5), 772-780 (2005)
PUBMED 15918889
COMMENT Contact: Wang S
National Key Laboratory of Crop Genetic Improvement
Hauzhong Agricultural University
Wuhan 430070, China
Tel: 86-27-87282044
Fax: 86-27-87287092
Email: shipingwang@hotmail.com
Seq primer: T7.
FEATURES Location/Qualifiers
source 1. .1781
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/strain="indica"
/cultivar="Minghui 63"
/db_xref="taxon:39946"
/clone="EI057O14, THREEB09, EI057O04, EI062D05, BI027I14, EI070G16, EI08"
/tissue_type="whole plant"
/dev_stage="whole-life-cycle"
/lab_host="E. coli DH10B"
/clone_lib="A normalized whole-life-cycle cDNA library of rice"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The library is constructed based on the strategy of satuation hybridization with genomic DNA using rice cultivar Minghui 63. This library consists of cDNA from 15 directionally cloned cDNA libraries constructed with different tissues from 9 developmental stages."
ORIGIN

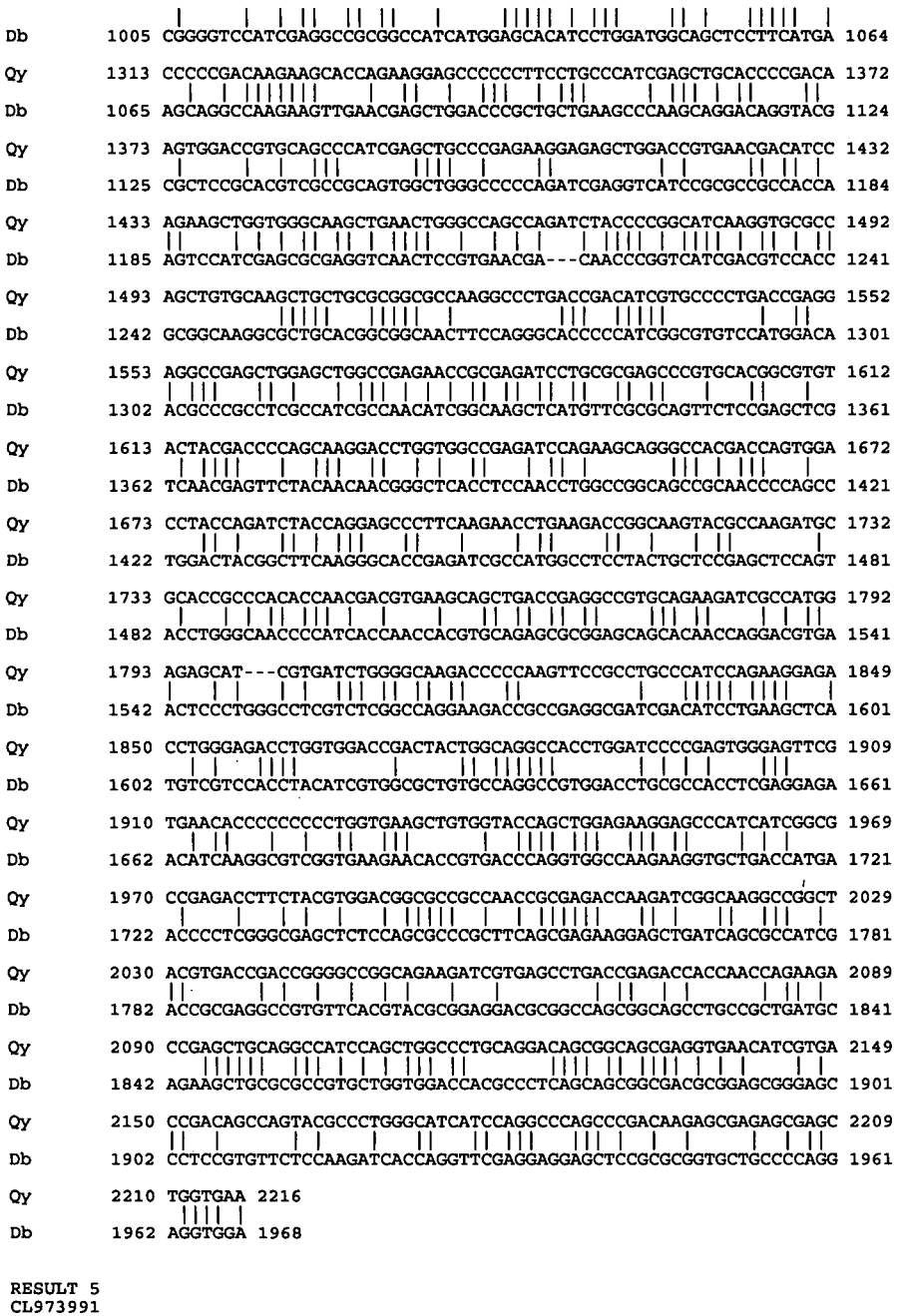
Query Match 3.7%; Score 90.2; DB 8; Length 1781;
Best Local Similarity 43.8%; Pred. No. 6.5e-06;
Matches 447; Conservative 0; Mismatches 568; Indels 6; Gaps 1;

```
Qy        231 GGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAA 290
Db        194 GGACGCGGTGCTCGACGCGTGCCTCGCCCAGGACCCCGACAGCAAGGTGGCTTGCGAGAC 253

Qy        291 CAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGC 350
Db        254 GTGCACCAAGACCAACATGGTGATGGTGTTCGGCGAGATCACCACCAAGGCCACCGTCGA 313

Qy        351 CGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAG 410
Db        314 CTACGAGAAGATCGTCCGCGACACCTGCCGCGGCATCGGCTTCGTGTCCGACGACGTCGG 373

Qy        411 CATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGT 470
Db        374 CCTCGACGCCGACCGCTGCAAGGTGCTCGTCAACATCGAGCAGCAGTCGCCCGACATCGC 433

Qy        471 GCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGG 530
Db        434 GCAGGGGGTGCACGGCCACTTCACCAAGCGCCCCGAGGAGATCGGCGCCGGCGACCAGGG 493

Qy        531 CTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCAT 590
Db        494 CCACATGTTCGGCTACGCCACCGACGAGACCCCCGAGCTGATGCCCCTCAGCCACGTCCT 553

Qy        591 CGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCA 650
Db        554 CGCCACCAAGCTCGGCGCCCGCCTCACCGAGGTCCGCAAGAACGGCACCTGCGCGTGGCT 613

Qy        651 GCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAA 710
Db        614 CAGGCCCGACGGCAAGACCCAGGTCACCGTCGAGTACCTCAACGACGCCGGCGCCATGGT 673

Qy        711 GCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGC 770
Db        674 CCCCGTCCGCGTCCACACCGTCCTCATCTTCACCCAGCACGACGAGACCGTCACCAACGA 733

Qy        771 CCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGA 830
Db        734 CGAGATCGCCGCC------GACCTCAAGGAGCACGTCATCAAGCCGGTCATCCCCGACAA 787

Qy        831 GAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAA 890
Db        788 GTACCTCGACGAGAAGACCATCTTCCACCTCAACCCCTCGGGCCGCTTCGTCATCGGCGG 847

Qy        891 GCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGG 950
Db        848 GCCCCACGGCGACGCCGGCCTCACCGGCCGCAAGATCATCATCGACACCTACGGCGGATG 907

Qy        951 CATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGA 1010
Db        908 GGGCGCGCACGGCGGCGGCGCCTTCTCCGGCAAGGACCCGACCAAGGTCGACCGGAGCGG 967
```

```
Qy       1011 CGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCC 1070
             ||||||| ||  |  ||     | | ||  | ||  |     |  |  |  |  |  |
Db        968 CGCCTACATCGCCAGGCAGGCCGCCAAGAGCATCGTCGCCAGCGGCCTCGCCCGCCGCTG 1027

Qy       1071 CAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTG 1130
             || |  |      |  | ||| |  |  |  | ||     |    ||  |   | ||
Db       1028 CATCGTGCAGGTGTCGTACGCCATCGGCGTCCCCGAGCCGCTCTCCGTGTTCGTCGACTC 1087

Qy       1131 GAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGC 1190
                | |||| |  ||  |||  | | |    ||  |||||| |  ||    |  ||
Db       1088 CTACGGCACCGGCAAGATCCCCGACAAGGAGATCCTCAAGATCGTCAAGGAGAACTTCGA 1147

Qy       1191 CCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGAT 1250
                 ||  |||| |||  |||  |  | |   |  |||  |   |||| || ||| |   ||
Db       1148 TTTCAGGCCCGGGATGATGACCATCAACCTCGACCTCAAGAGGGGCGGCAACCGGTTCAT 1207

Qy       1251 C 1251
              |
Db       1208 C 1208
```

```
RESULT 4
AY103647
LOCUS       AY103647                2598 bp    mRNA    linear   HTC 18-FEB-2005
DEFINITION  Zea mays PCO142084 mRNA sequence.
ACCESSION   AY103647
VERSION     AY103647.1  GI:21206725
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2598)
  AUTHORS   Gardiner,J., Schroeder,S., Polacco,M.L., Sanchez-Villeda,H.,
            Fang,Z., Morgante,M., Landewe,T., Fengler,K., Useche,F.,
            Hanafey,M., Tingey,S., Chou,H., Wing,R., Soderlund,C. and Coe,E.H.
            Jr.
  TITLE     Anchoring 9,371 maize expressed sequence tagged unigenes to the
            bacterial artificial chromosome contig map by two-dimensional
            overgo hybridization
  JOURNAL   Plant Physiol. 134 (4), 1317-1326 (2004)
   PUBMED   15020742
REFERENCE   2  (bases 1 to 2598)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   3  (bases 1 to 2598)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .2598
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeGDB:638378"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN
```

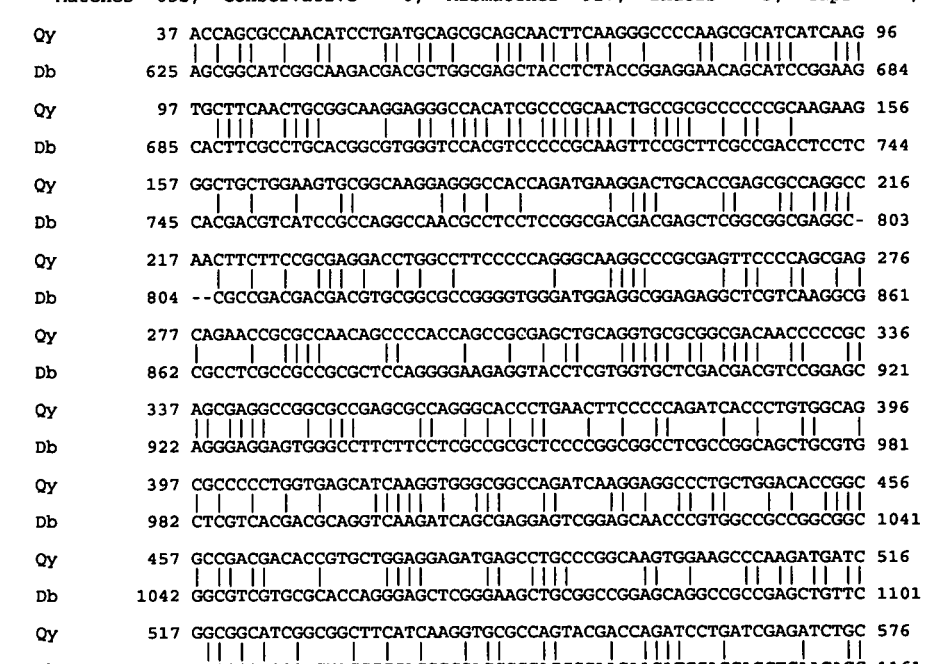
```
  Query Match             3.7%;  Score 90.2;  DB 4;  Length 2598;
  Best Local Similarity   42.2%;  Pred. No. 6.7e-06;
  Matches  814;  Conservative    0;  Mismatches 1098;  Indels   15;  Gaps    5;

Qy        296 CCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGC 355
              |||||| |  | | || | |   |  ||| |  |  ||      |   | |||| |
Db         51 CCACCACCTCCCACCTCCTGCCACCGCGCGGCCCCAACCAACCACACCGCAGCGCAGCAA 110

Qy        356 GCCAGGGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCA 415
              |||||  |      | ||    ||| |  || | |  |  |    |  ||| |
Db        111 TGGCGGGCAACGGCGCCATCGTGGAGAGCGACCCGCTGAACTGGGGCGCGGCGGCGGCGG 170

Qy        416 AGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGG 475
              || ||| |||       |  ||| |   ||  |  ||  |  ||  | | |     || |
Db        171 AGCTGGCCGGGAGCCACCTGGACGAGGTGAAGCGCATGGTGGCGCAGGCCCGGCAGCCCG 230

Qy        476 AGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCA 535
              ||  | || |  |     ||       |     | | | || | ||| |  |  ||
Db        231 TGGTCAAGATCGAGGGTTCCACCCTCCGCGTCGGCCAGGTGGCCGCCGTCGCCTCCGCCA 290

Qy        536 TCAAGGTGCGCCAGTACGACCAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCA 595
              | | |  |    || |      ||  ||||      | ||      | ||| | ||
Db        291 AGGACGCGTCCGGCGTCGCCGTCGAGCTCGACGAGGAGGCCCGCCCCCGCGTCAAGGCCA 350

Qy        596 CCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGG 655
              |     |    || |  | | |  |  |  |||| || |      |     |
Db        351 GCAGCGAGTGGATCCTCGACTGCATCGCCCACGGCGGCGACATCTACGGCGTCACCACCG 410

Qy        656 GCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCG 715
              ||| |  |  | || || |  | || |||  ||      |  |  || |  |||
Db        411 GCTTCGGCGGCACCTCCCACCGCCGCACCAAGGACGGGCCCGCGCTCCAGGTCGAGCTGC 470

Qy        716 GCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGA 775
              || | |  |  |  |  |  |  |  ||  | ||| | | |  |  | |  |
Db        471 TCAGGCATCTCAACGCCGGAATCTTCGGCACCGGCAGCGACGGGCACACGCTGCCGTCGG 530

Qy        776 CCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACC 835
              | || | ||| || |||| |  |     ||| | |  |      ||| |    |
Db        531 AGGTCACCCGCGCGGCGATGCTGGTGCGCATCAACACCCTCCTCCAGGGCTACTCCGGCA 590

Qy        836 CCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGG 895
              | | |    | |   |||||| |||| |  || ||||        ||
Db        591 TCCGCTTCGAGATCCTCGAGGCCATCACGAAGCTGCTCAACACCGGTGTCAGCCCCTGCC 650

Qy        896 TGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCC 955
              ||    |||| |     | ||       |  |||  |  |  |    | |
Db        651 TGCCGCTCCGGGGCACCATCACCGCGTCGGGCGACCTGGTCCCGCTCTCCTACATCGCCG 710

Qy        956 CCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCT 1015
              || |  | ||||     ||   | ||  ||| ||||    | ||||  |  |  | |
Db        711 GCCTCATCACGGGCCGCCCCAACGCGCAGGCCGTCACCG---TCGACGGGAGGAAGGTGG 767

Qy       1016 ACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCA 1075
              || |  || | |  |  |    | |  |  | ||  |  |  |||||   ||  |  ||
Db        768 ACGCCGCCGAGGCGTTCAAGATCGCCGGCATCGAGGGCGGCTTCTTCAAGCTCAACCCCA 827

Qy       1076 TCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGG 1135
              |  |  |    ||  || || ||                |  || |  |  |  |
Db        828 AGGAGGGCCTCGCCATCGTCAACGGCACGTCCGTGGGCTCCGCGCTCGCGGCCACCGTGA 887

Qy       1136 GCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCA 1195
              |  |  || | ||| |  |  | ||| |  || ||||   | ||  |  |  | ||
Db        888 TGTACGACGCCAACGTCCTGGCCGTCCTGTCGGAGGTCCT---GTCCGCCGTCTTCTGCG 944

Qy       1196 ACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCC 1255
              |  |  || ||  |    | ||  | |      || ||  |  | | ||| ||  |  ||
Db        945 AGGTCATGAACGGCAAGCCCGAGTACACGGACCACCTGACCCACAAGCTGAAGCACCACC 1004

Qy       1256 AGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG---CGCTGGGGCTTCACCA 1312
```

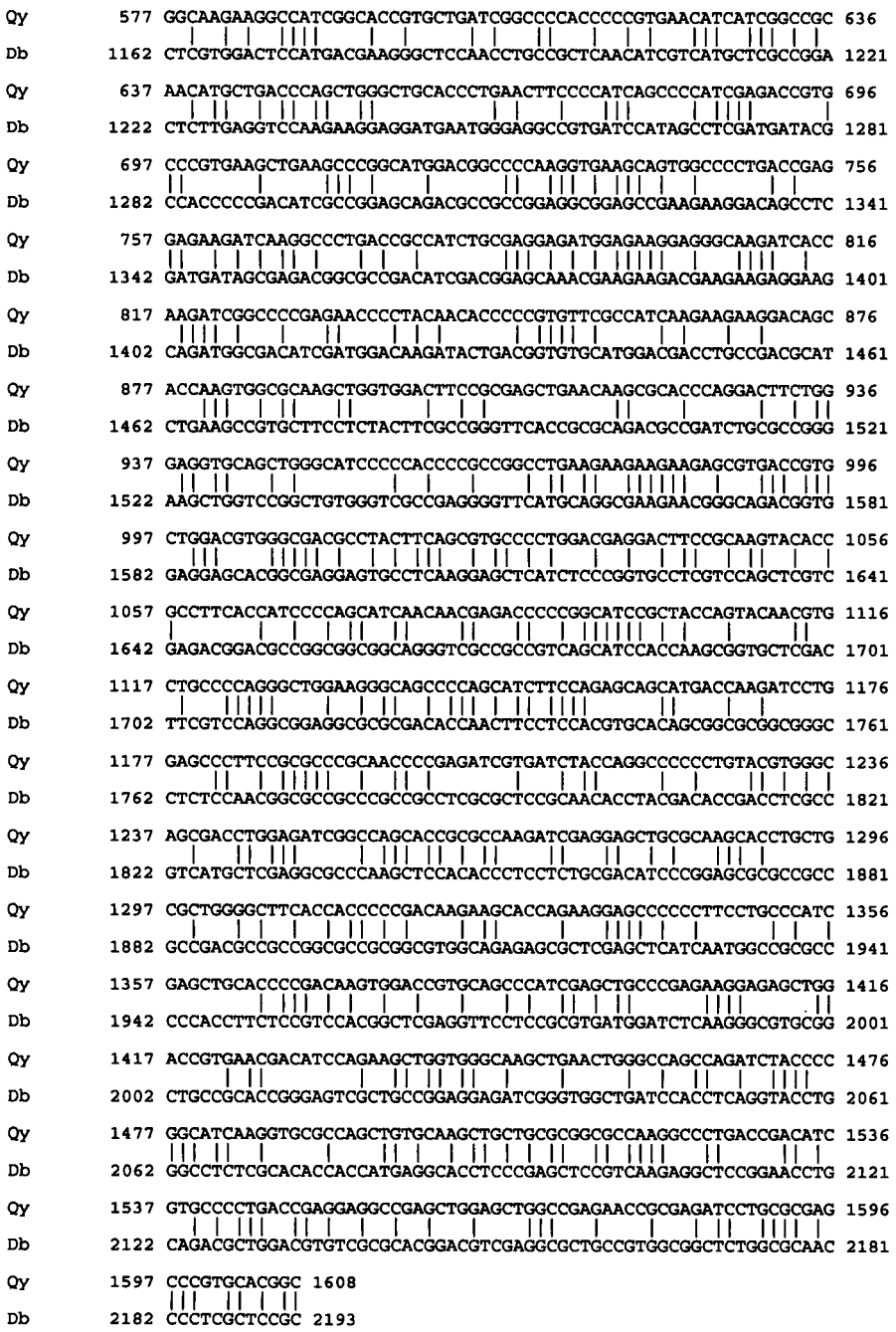


RESULT 5
CL973991
LOCUS CL973991 3069 bp DNA linear GSS 21-SEP-2004
DEFINITION OsIFCC025416 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.

ACCESSION CL973991
VERSION CL973991.1 GI:52402507
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 3069)
AUTHORS Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES Location/Qualifiers
source 1. .3069
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Query Match 3.6%; Score 88.8; DB 10; Length 3069;
Best Local Similarity 41.5%; Pred. No. 1.2e-05;
Matches 652; Conservative 0; Mismatches 917; Indels 3; Gaps 1;





```
RESULT 6
CL961989
LOCUS       CL961989               1398 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC006991 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL961989
VERSION     CL961989.1  GI:52378720
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1398)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .1398
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

  Query Match             3.5%;  Score 86.6;  DB 10;  Length 1398;
  Best Local Similarity   45.9%;  Pred. No. 2.8e-05;
  Matches  508;  Conservative    0;  Mismatches  569;  Indels   30;  Gaps    5;

Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              ||||| | | | |||||||| | |  |  || |||| ||  |  ||
Db        166 GTCGAGAGGCACGTCGGCGAGGCCGACATCGTCTTCGTCTCCGTGAACACCCCCACCAAG 225

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
                | ||  | ||   | |  |||| |  |   |  |  | |||| |  |  | ||
Db        226 GCCCGCGGCCTCGGCGCCGGCAAGGCCGCCG---ACCTCACCTACTGGGAGAGCGCGGCG 282

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              |  ||   |  |  |     |  ||  ||||   | |  |  |||||  |  |  |
Db        283 CGGATGATCGCCGCCGTGGCGACGTCCGACAAGGTCGTCGTCGAGAAGTCCACCGTGCCG 342

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
              | | |       | ||| |   |    ||  ||  | | ||  |  ||||||  | | | |
Db        343 GTCAAGACCGCGGAGGCCATCGAGAAGATCCTCGACCACAACGGCCGCGACGGCGTCGGC 402

Qy        241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACC 300
              ||||     |    ||  ||||||| |  ||||   | |  |  |  |  ||     |
Db        403 TTCCAGATCCTCTCCAACCCGGAGTTCCTCGCCGAGGGCACCGCCATCCGCGACCTGCTC 462

Qy        301 AGCCGCGAGCTGCAGGTGCGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAG 360
                || |||          |||| |  || |  | || ||||| ||| ||| | ||||
Db        463 GCCCCCGA---------CCGCGTCCTCATCGGCGGCCGCGAGACCGCCGCGGGGCGCGCC 513

Qy        361 GGCACCCTGAACTTCCCCCAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTG 420
              | |  || |    ||    |  |    |   |  |  ||||  || ||| |    |
Db        514 GCCGTCCAGGCGCTCAAGGACGTGTACGCCCGATGGGTCCCCGAGGAGAGGATC---CTC 570

Qy        421 GGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAG 480
              | | | ||  |   |||  ||||  ||    ||||| | ||  |  ||||    ||
Db        571 ACCACCAACCTCTGGTCCGCCGAGCTGTCCAAGCTCGCCGCCAACGCGTTCCTGGCCCAG 630
```



```
Qy        577 GGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGC 636
Db       1162 CTCGTGGACTCCATGACGAAGGGCTCCAACCTGCCGCTCAACATCGTCATGCTCGCCGGA 1221

Qy        637 AACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTG 696
Db       1222 CTCTTGAGGTCCAAGAAGGAGGATGAATGGGAGGCCGTGATCCATAGCCTCGATGATACG 1281

Qy        697 CCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAG 756
Db       1282 CCACCCCCGACATCGCCGGAGCAGACGCCGCCGGAGGCGGAGCCGAAGAAGGACAGCCTC 1341

Qy        757 GAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACC 816
Db       1342 GATGATAGCGAGACGGCGCCGACATCGACGGAGCAAACGAAGAAGACGAAGAAGAGGAAG 1401

Qy        817 AAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGC 876
Db       1402 CAGATGGCGACATCGATGGACAAGATACTGACGGTGTGCATGGACGACCTGCCGACGCAT 1461

Qy        877 ACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG 936
Db       1462 CTGAAGCCGTGCTTCCTCTACTTCGCCGGGTTCACCGCGCAGACGCCGATCTGCGCCGGG 1521

Qy        937 GAGGTGCAGCTGGGCATCCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTG 996
Db       1522 AAGCTGGTCCGGCTGTGGGTCGCCGAGGGGTTCATGCAGGCGAAGAACGGGCAGACGGTG 1581

Qy        997 CTGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACC 1056
Db       1582 GAGGAGCACGGCGAGGAGTGCCTCAAGGAGCTCATCTCCCGGTGCCTCGTCCAGCTCGTC 1641

Qy       1057 GCCTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTG 1116
Db       1642 GAGACGGACGCCGGCGGCGGCAGGGTCGCCGCCGTCAGCATCCACCAAGCGGTGCTCGAC 1701

Qy       1117 CTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTG 1176
Db       1702 TTCGTCCAGGCGGAGGCGCGCGACACCAACTTCCTCCACGTGCACAGCGGCGCGGCGGGC 1761

Qy       1177 GAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGC 1236
Db       1762 CTCTCCAACGGCGCCGCCCGCCGCCTCGCGCTCCGCAACACCTACGACACCGACCTCGCC 1821

Qy       1237 AGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTG 1296
Db       1822 GTCATGCTCGAGGCGCCCAAGCTCCACACCCTCCTCTGCGACATCCCGGAGCGCGCCGCC 1881

Qy       1297 CGCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATC 1356
Db       1882 GCCGACGCCGCCGGCGCCGCGGCGTGGCAGAGAGCGCTCGAGCTCATCAATGGCCGCGCC 1941

Qy       1357 GAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGG 1416
Db       1942 CCCACCTTCTCCGTCCACGGCTCGAGGTTCCTCCGCGTGATGGATCTCAAGGGCGTGCGG 2001

Qy       1417 ACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCC 1476
Db       2002 CTGCCGCACCGGGAGTCGCTGCCGGAGGAGATCGGGTGGCTGATCCACCTCAGGTACCTG 2061

Qy       1477 GGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCTGACCGACATC 1536
Db       2062 GGCCTCTCGCACACCACCATGAGGCACCTCCCGAGCTCCGTCAAGAGGCTCCGGAACCTG 2121

Qy       1537 GTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGCGAG 1596
Db       2122 CAGACGCTGGACGTGTCGCGCACGGACGTCGAGGCGCTGCCGTGGCGGCTCTGGCGCAAC 2181

Qy       1597 CCCGTGCACGGC 1608
Db       2182 CCCTCGCTCCGC 2193
```

```
Qy        481 ATGAGCCTGCCCGGCAAGTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG 540
Db        631 AGGATCTCGTCCGTCAACGCCATGTCCGCGCTCTGCGAGGCCACCGGCGCCGACGTCGCC 690

Qy        541 GTGCGCCAGTACGACCAGATCCTGATCGA---GATCTGCGGCAAGAAGGCCATCGGCACC 597
Db        691 GAGGTCGCCTACGCCGTGGGCAAGGACTCCAGGATCGGCGCCAAGTTTCTCAACGCCAGC 750

Qy        598 GTGCTGATCGGCCCCACCCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGC 657
Db        751 GTGGGATTCGGTGGCTCCTGCTTCCAGAAAGACATCCTCAACCTGCTG------------ 798

Qy        658 TGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGC 717
Db        799 TACATCTGCGAGTGCAATGGCCTCCCCGAGGTGGCCAACTACTGGAAGCAGGTTATCAAG 858

Qy        718 ATGGACGGCCCCAAGGTGAAGCAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACC 777
Db        859 ATCAACGACTACCAGAAGAGCCGGTTCGTGAACCGCGTGGTGTCGTCCATGTTCAACACC 918

Qy        778 GCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCC 837
Db        919 GTGGCCGGCAAGAAGATCGCCGTGCTGGGCTTCGCCTTCAAGAAGGACACCGGCGACACC 978

Qy        838 TACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTG 897
Db        979 AGGGAGACGCCCGCCATCGACGTCTGCAAGGGGCTGATCGGCGACAAGGCCAAGGTGAGC 1038

Qy        898 GACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCC 957
Db       1039 ATCTACGACCCCCAGGTGACGGAGGACCAGGTCCAGCGCGACCTCGCCATGAGCAAGTTC 1098

Qy        958 CACCCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTAC 1017
Db       1099 GATTGGGATCACCCGGTCCATCTCCAGCCGATGAGCCCCACGGCCATCAAGCAGGTGTCC 1158

Qy       1018 TTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCATC 1077
Db       1159 GTCGCCTGGGACGCCTACGAGGCGGCCAGGGACGCCCACGGCGTCTGCATCCTCACCGAG 1218

Qy       1078 AACAACGAGACCCCCGGCATCCGCTAC 1104
Db       1219 TGGGACGAGTTCAGGAGCCTCGACTAC 1245
```

RESULT 7
CZ247380
LOCUS CZ247380 743 bp DNA linear GSS 11-FEB-2005
DEFINITION AIAA-aaf35a12.b1 Ancylostoma caninum whole genome shotgun library (AIAAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION CZ247380
VERSION CZ247380.1 GI:59632821
KEYWORDS GSS.
SOURCE Ancylostoma caninum (dog hookworm)
ORGANISM Ancylostoma caninum
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida; Ancylostomatoidea; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
REFERENCE 1 (bases 1 to 743)
AUTHORS Mitreva,M., McCarter,J.P., Pape,D., Ritter,E., Tsagareishvili,R., Ronko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Messina,D., Waterston,R.H., Clifton,S.W. and Wilson,R.
TITLE Genome Survey sequences from the parasitic nematode Ancylostoma caninum
JOURNAL Unpublished (2004)
COMMENT Contact: Mitreva M
Washington University in St. Louis
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@watson.wustl.edu
Genomic DNA provided by John Hawdon (mtmjmh@gwumc.edu) DNA sequenced by Washington University Genome Sequencing Center
Class: shotgun.

FEATURES Location/Qualifiers
source 1. .743
/organism="Ancylostoma caninum"
/mol_type="genomic DNA"
/strain="Baltimore"
/db_xref="taxon:29170"
/dev_stage="Adult"
/lab_host="GS10"
/clone_lib="Ancylostoma caninum whole genome shotgun library (AIAAGSS 001)"
/note="Vector: pOTW13; Site_1: BstX1; Site_2: BstX1; Ancylostoma caninum genomic DNA was randomly sheared, end-repaired and size fractioned to enrich for 2-4 kb fragments. Genomic DNA was provided by John Hawdon (mtmjmh@gwumc.edu) at George Washington University. Sequencing by Washington University Genome Sequencing Center, St. Louis, MO."

ORIGIN

Query Match 3.5%; Score 86; DB 10; Length 743;
Best Local Similarity 45.1%; Pred. No. 3.5e-05;
Matches 320; Conservative 0; Mismatches 390; Indels 0; Gaps 0;

```
Qy        777 CGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCC 836
Db          5 CGACAACAACAACAACAACAACAACGACAACAACAACAACAACAACAACGTTAACAACAA 64

Qy        837 CTACAACACCCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT 896
Db         65 CGACAACAACAACGACAACGACAACAACGACTTCAACAACAACGACAACAACAACGACAA 124

Qy        897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC 956
Db        125 CAACAACAACGACGACAACAACGACAACAACGACAACAACAACGACAACAACAACAACAA 184

Qy        957 CCACCCCGCCGGCCTGAAGAAGAAGAAGAAGAGCGTGACCGTGCTGGACGTGGGCGACGCCTA 1016
Db        185 CAACAACAACAACAACGACAACAACGACAACAACAACGACAACAACAACAACGACGACGA 244

Qy       1017 CTTCAGCGTGCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCCAGCAT 1076
Db        245 CAACAACGACAACAACGACAACAACAACAACAACAACGACAACAACAACGACAACAA 304

Qy       1077 CAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGGGCTGGAAGGG 1136
Db        305 CAACAACGACGACGACAACAACGACAACAACGACAACGACAACAACAACAACAACAA 364

Qy       1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAA 1196
Db        365 CGACAACAACAACGACAACAACAACAACGACGACAACAACGACAACGACAACGACAACAA 424

Qy       1197 CCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGCGACCTGGAGATCGGCCA 1256
Db        425 CAACGACAGCAACAACGACAACGACAACAACAACAACGACAACAACGACAACAACAA 484

Qy       1257 GCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCACCCC 1316
Db        485 CAACAACGACAACAACAACGACAACAACAACGACAACAACAACAACAACAACGACAA 544

Qy       1317 CGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGCCCATCGAGCTGCACCCCGACAAGTG 1376
Db        545 CGACAACGACAACGACTACGACAACAACAACGACAACAACAACATCGACGACTACAACAA 604

Qy       1377 GACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAA 1436
Db        605 CAACGACAACAACAACGACGACAACTACAACAACGACAACTACTACAACGACAACAACTA 664

Qy       1437 GCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGG 1486
Db        665 CAACGACAACGACAACAACAACGACGACTACTACAACGACGACAACTACG 714
```

RESULT 8
BM321451
LOCUS BM321451 951 bp mRNA linear EST 03-JAN-2002
DEFINITION rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC 3.3.1.1), mRNA sequence.
ACCESSION BM321451
VERSION BM321451.1 GI:18055857
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 951)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 951 Std Error: 0.00
POLYA=No.
FEATURES Location/Qualifiers
source 1. .951
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"
ORIGIN

Query Match 3.5%; Score 85.6; DB 3; Length 951;
Best Local Similarity 46.0%; Pred. No. 4.2e-05;
Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;

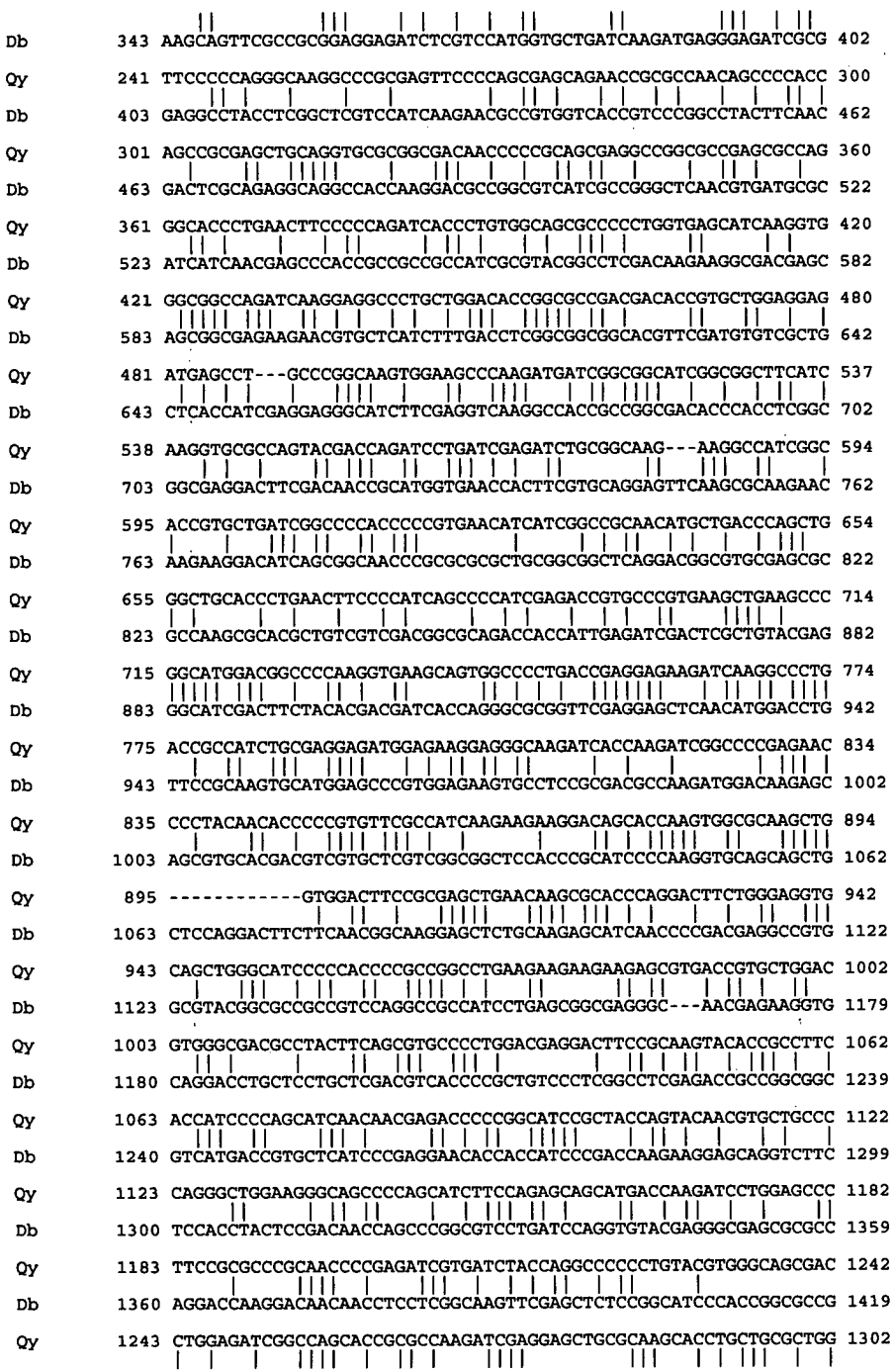
```
Qy       1722 CGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAA 1781
             ||||||||| |||     |     ||||  | |   |  || |  | | || |
Db        219 CGCCAAGATCCGCTGGTGCTCGTGCAACATCTTCTCGACGCAGGACCACGCGGCCGCCGC 278

Qy       1782 GATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAAGTTCCGCCTGCCCATCCA 1841
              |||||   ||   || ||  | |||  | |  ||    | ||  || || |     |
Db        279 CATCGCGCAGGCCGGCGTCTCGGTCTTCGCCTGGAAGGGCGAGAACCTCCAGGAGTACTG 338

Qy       1842 GAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTG 1901
              | ||   |||||| || ||  |||   || |  ||  |||| | |   ||||  ||
Db        339 GGAGTGCACCTGGAAGGCCCTGTGCTTCGGCCCCTACCAGGGCCCTCAGATCATCGTCGA 398

Qy       1902 GGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCAT 1961
               ||   ||   || |  | |   || |  |   | ||| |  |   ||| |  | ||||
Db        399 CGACGGCGGTGACGCGACTCTGATGATCCCACAAGGGGTTCGCGGCCGAGGACAACCCCAA 458

Qy       1962 CATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAA 2021
                |   |  ||   |   | | || |  | |||| || | | | |  | ||   |
Db        459 GCTGCTGGAGGACGACGAGGGCCTCGAGGAGGTCGCCTGCCTCAACAACGTGCTCAAGCA 518

Qy       2022 GGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAGCCTGACCGAGACCACCAA 2081
              || |   | | |   | |  | |||| |||||| |   |  || |       ||
Db        519 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCCTCCCCGAGATCCGCGGTGTCAG 578

Qy       2082 CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGGTGAA 2141
              | || ||||    |     |  ||   ||||  || || | || ||| | || || ||
Db        579 CGAGGAGACGACGACTGGCGTNATGAGGCTGTACCAGCTGCACCGCGACGGCAAGCTGCT 638

Qy       2142 CATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAGCGA 2201
               ||  | ||| || |   ||| |     ||| | ||  | ||  ||||||  | | |
Db        639 GTTCCCGGCCGTCAACGTCAACGAC---TCTNTCACCAAGAGCAAGTTCGACAACATCTA 695
```

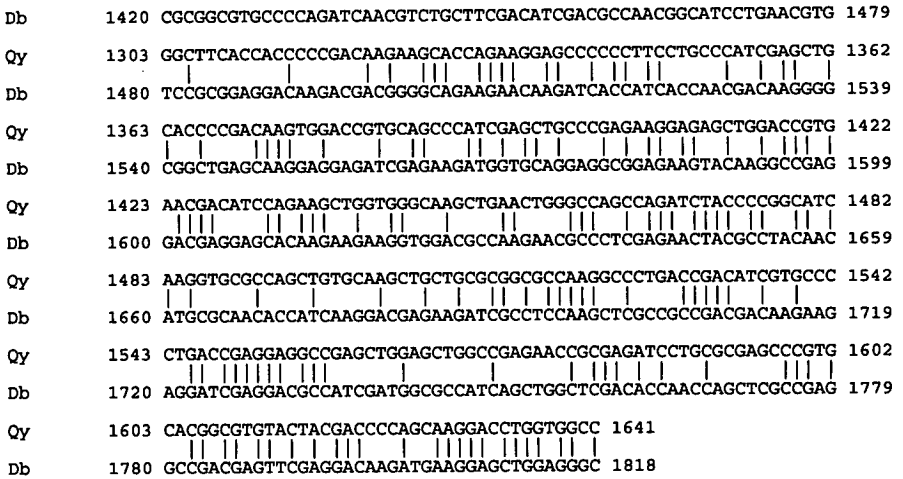
```
Qy       2202 GAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCTGAG 2261
               ||  |     |  |||  ||     |   |   |  ||  | ||  |  |
Db        696 CGGCTGCCGCCACTCGCTCATCGACGGCATCAAGCGCGGCACCGACGTGATGCTCGGCGG 755

Qy       2262 CTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2321
              |  |||  ||| |  |  ||||  ||||| |   |   ||  | |     |  |
Db        756 CAAGGTCGCCGTCGTCGCGGGCTACGGCGACGTGGGCAAGGGCTGCGCCGAGTCGCTGCG 815

Qy       2322 GGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATGGCGGCATCGTGATCTACCAGTA 2381
              |||    ||         |  || ||   | |||||  |  |  || | |   ||
Db        816 CGGCCAGGGCTGCCGCGTCATCGTGACGGAGATCGACCCCATCTGCGCGCTGCAGGCGTC 875

Qy       2382 CATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGA 2423
              |||| || | |  | ||   ||     ||   |  || ||||
Db        876 GATGGCCGGCTTCGAGGTCAACACGCTCGAGGCGGGGCTCGA 917
```

RESULT 9
CL971508
LOCUS CL971508 1941 bp DNA linear GSS 21-SEP-2004
DEFINITION OsIFCC021485 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL971508
VERSION CL971508.1 GI:52397596
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 1941)
AUTHORS Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES Location/Qualifiers
source 1. .1941
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Query Match 3.4%; Score 84.2; DB 10; Length 1941;
Best Local Similarity 42.5%; Pred. No. 7.8e-05;
Matches 705; Conservative 0; Mismatches 933; Indels 21; Gaps 4;

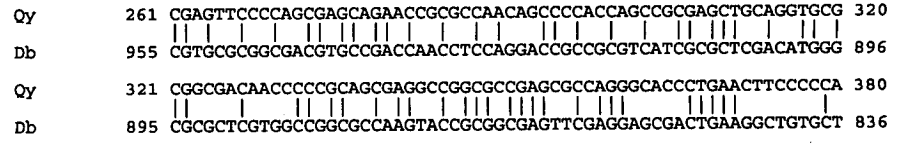
```
Qy          1 GTCGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAACATCCTGATGCAG 60
              | || |||||  |      | |||||| ||   ||| || || | |      || || |
Db        163 GACGCCGCCAAGAACCAGGTCGCCATGAACCCAACCAACACCGTCTTCGATGCGAAGCGG 222

Qy         61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGC 120
                |  | | || | |  | |  || |  ||  |  |     | | |
Db        223 TTGATTGGTAGGAGGTTCTCCGACCCGTCCGTGCAGAGTGACATGAAGCTATGGCCGTTC 282

Qy        121 CACATCGCCCGCAACTGCCGCGCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAG 180
              |  | |  |    | | ||| |   | | ||  | |  | | |    ||| |||||
Db        283 AAGGTGGTGCCTGGCCCCGGCGACAAGCCGATGATCGTCGTCCAGTACAAGGGCGAGGAG 342

Qy        181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240
```

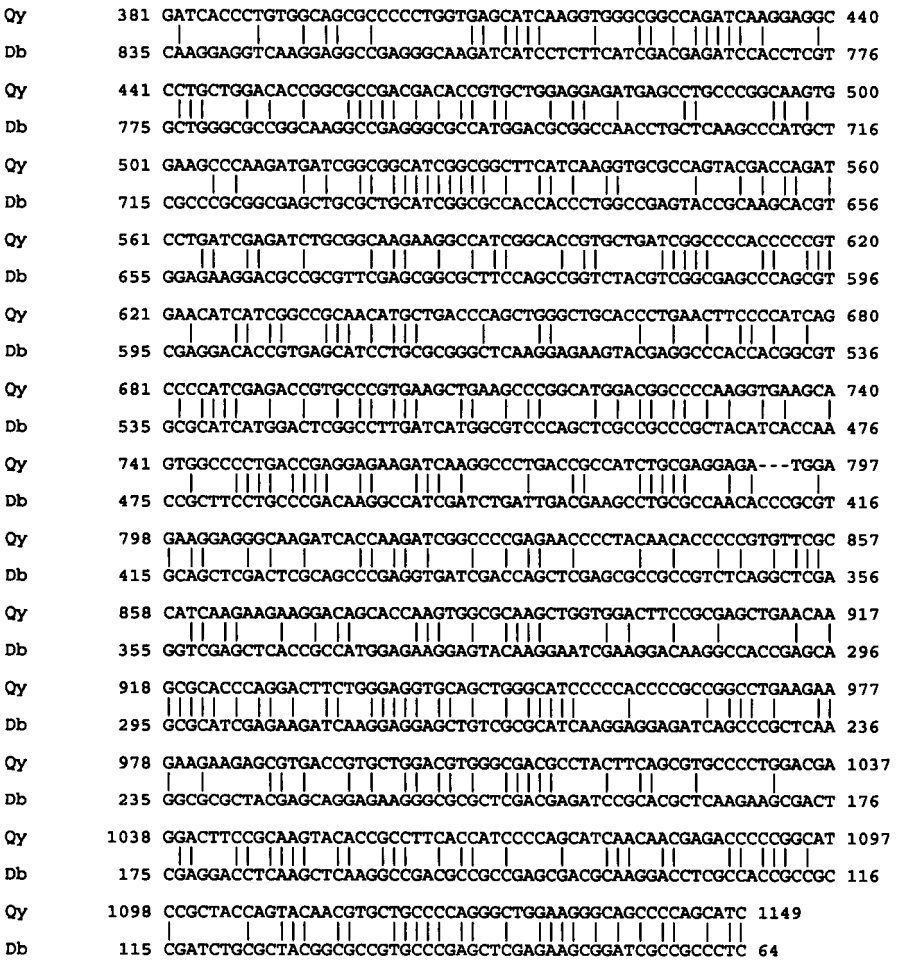




RESULT 10
CW922203/c
LOCUS CW922203 1060 bp DNA linear GSS 24-FEB-2005
DEFINITION EDCAR29TR A. castellanii, 6-8 kb library from total genomic DNA Acanthamoeba castellanii genomic clone EDCAR29, genomic survey sequence.
ACCESSION CW922203
VERSION CW922203.1 GI:60228146
KEYWORDS GSS.
SOURCE Acanthamoeba castellanii
ORGANISM Acanthamoeba castellanii
Eukaryota; Acanthamoebidae; Acanthamoeba.
REFERENCE 1 (bases 1 to 1060)
AUTHORS Anderson,I.J. and Loftus,B.J.
TITLE Gene discovery in the Acanthamoeba castellanii genome
JOURNAL Unpublished (2004)
COMMENT Contact: Iain Anderson
The Institute for Genomic Research (TIGR; www.tigr.org)
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-795-7949
Fax: 301-838-0208
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .1060
/organism="Acanthamoeba castellanii"
/mol_type="genomic DNA"
/strain="Neff"
/db_xref="taxon:5755"
/clone="EDCAR29"
/clone_lib="A. castellanii, 6-8 kb library from total genomic DNA"
/note="Vector: pHOS2"
ORIGIN

Query Match 3.3%; Score 80.8; DB 10; Length 1060;
Best Local Similarity 43.9%; Pred. No. 0.0003;
Matches 392; Conservative 0; Mismatches 497; Indels 3; Gaps 1;

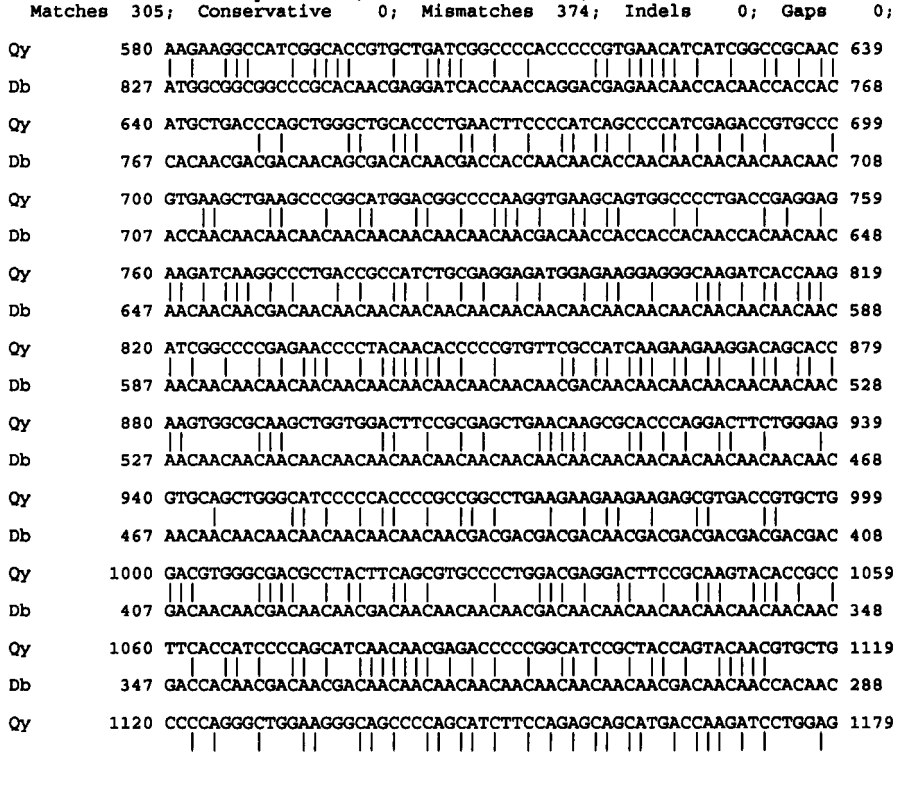




RESULT 11
CK159167/c
LOCUS CK159167 869 bp mRNA linear EST 05-DEC-2003
DEFINITION FGAS040564 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA, mRNA sequence.
ACCESSION CK159167
VERSION CK159167.1 GI:38985053
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 869)
AUTHORS Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D., Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A., Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D., Penniket,C., Roach,J.L. and Sarhan,F.
TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL Unpublished (2003)

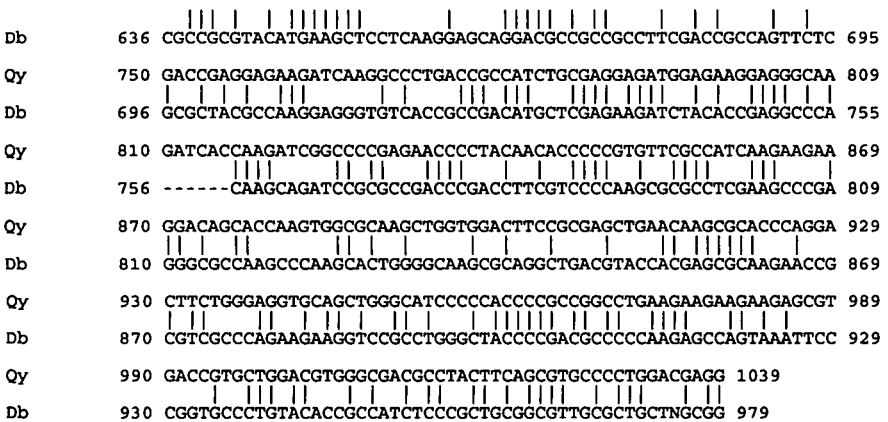
COMMENT Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas_ests@cs.usask.ca
This sequence is the direct result of the Base calling software Phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [128,636].
Plate: TaLt537 row: N column: 23.
FEATURES Location/Qualifiers
source 1. .869
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DH5 alpha"
/clone_lib="Triticum aestivum FGAS: TaLt5"
/note="Organ: Crown; Vector: pGEM-T; SSH (suppression subtractive hybridization) cDNA library from genotype PI178383 cold hardened at 2 C for 21 days and 49 days (equal amount of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 H)(driver). Modified Smart cDNA (Clontech)priming and non-directional cloning"
ORIGIN

Query Match 3.3%; Score 80.6; DB 7; Length 869;
Best Local Similarity 44.9%; Pred. No. 0.00033;
Matches 305; Conservative 0; Mismatches 374; Indels 0; Gaps 0;



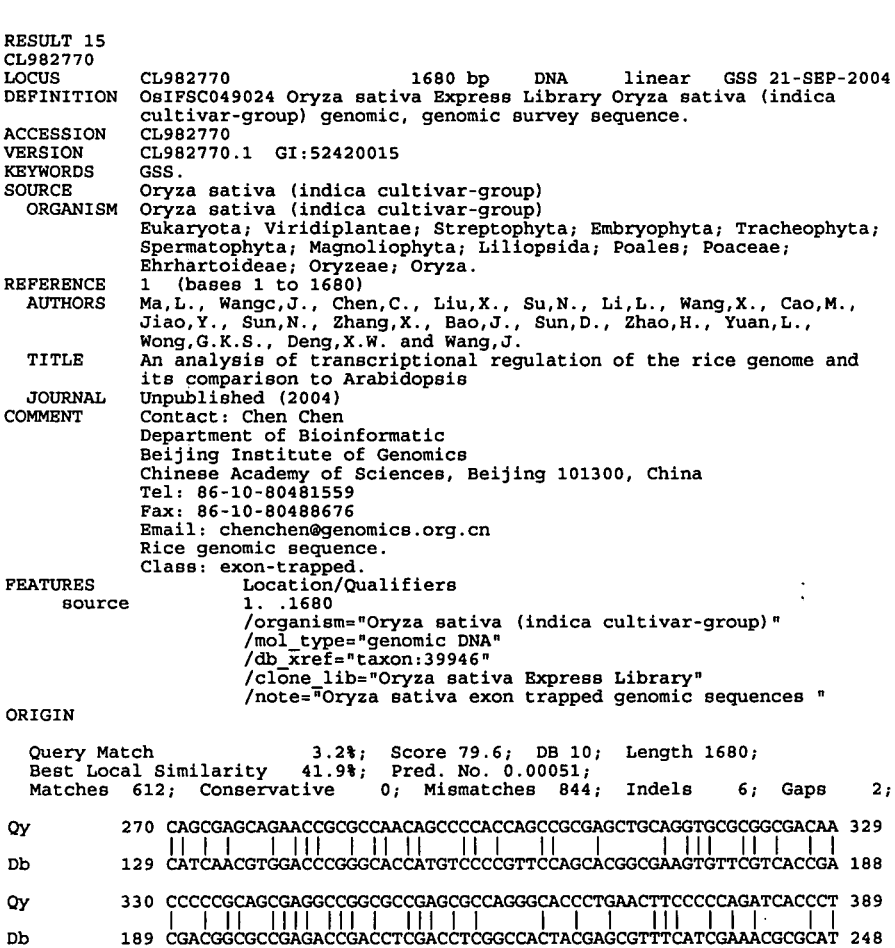
```
Db        287 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 228

Qy       1180 CCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAGGCCCCCCTGTACGTGGGCAGC 1239

Db        227 AACAACAACAACAACAACGACGACAACGACGTCAACGATAACAACGACAACAACGACAAC 168

Qy       1240 GACCTGGAGATCGGCCAGC 1258

Db        167 AACGGCAACAACAACAAAC 149


RESULT 12
CZ216254
LOCUS       CZ216254                892 bp    DNA     linear   GSS 10-FEB-2005
DEFINITION  AIAA-aaf23g02.b1 Ancylostoma caninum whole genome shotgun library
            (AIAAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION   CZ216254
VERSION     CZ216254.1  GI:59229909
KEYWORDS    GSS.
SOURCE      Ancylostoma caninum (dog hookworm)
  ORGANISM  Ancylostoma caninum
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
            Ancylostomatoidea; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
REFERENCE   1  (bases 1 to 892)
  AUTHORS   Mitreva,M., McCarter,J.P., Pape,D., Ritter,E., Tsagareishvili,R.,
            Ronko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Messina,D.,
            Waterston,R.H., Clifton,S.W. and Wilson,R.
  TITLE     Genome Survey sequences from the parasitic nematode Ancylostoma
            caninum
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Mitreva M
            Washington University in St. Louis
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: nematode@watson.wustl.edu
            Genomic DNA provided by John Hawdon (mtmjmh@gwumc.edu) DNA
            sequenced by Washington University Genome Sequencing Center
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .892
                     /organism="Ancylostoma caninum"
                     /mol_type="genomic DNA"
                     /strain="Baltimore"
                     /db_xref="taxon:29170"
                     /dev_stage="Adult"
                     /lab_host="GS10"
                     /clone_lib="Ancylostoma caninum whole genome shotgun
                     library (AIAAGSS 001)"
                     /note="Vector: pOTW13; Site_1: BstX1; Site_2: BstX1;
                     Ancylostoma caninum genomic DNA was randomly sheared,
                     end-repaired and size fractioned to enrich for 2-4 kb
                     fragments. Genomic DNA was provided by John Hawdon
                     (mtmjmh@gwumc.edu) at George Washington University.
                     Sequencing by Washington University Genome Sequencing
                     Center, St. Louis, MO."
ORIGIN

  Query Match             3.3%;  Score 80.6;  DB 10;  Length 892;
  Best Local Similarity   46.5%;  Pred. No. 0.00033;
  Matches  334;  Conservative    0;  Mismatches  379;  Indels    6;  Gaps    2;

Qy       1644 GATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAA 1703

Db          5 GATTCAGCTCGAGCACCTCCAGCAGCACGTCAACCAGCACGTCAACCACCACCTCGACGT 64

Qy       1704 CCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCT 1763

Db         65 CGACGACACCAAGCACAACCACATCCACAACCACTTCGACGACACCGACGACTACTAGCT 124

Qy       1764 GACCGAGGCCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCCAA 1823

Db        125 CTAGCACGTCAAGCAGCACGTCAACCAGCACGTCAACCAGCACATCGACGTCGACGACAC 184

Qy       1824 GTTCCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGC 1883

Db        185 CAAGCACAACTACGTCCACCACCACTTCGACGACACCGACGACTACTAGCTCTAGCACGT 244

Qy       1884 CACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCTGGTGAAGCTGTGGTACCA 1943

Db        245 CAAGCAGCACGTCAACCAGCACGTCAACCAGCACCTCGACGTCGACGACACCAAGCCCAA 304

Qy       1944 GCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCG 2003

Db        305 CCACGTCCACAACCACTTCGACGACACCGACGACTACTACCTCGAGTACCTCGACCAGCA 364

Qy       2004 CGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGATCGTGAG 2063

Db        365 CGTCAACCAGCACCTCGACGTCGACGACACCAAGCACAACCACGTCCACAACCACTTCGA 424

Qy       2064 CCTGACCGAGACCACCAA---CCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCA 2120

Db        425 CGACACCGACGACTACTAGCTCGAGCACGTCAAGCAGCACGTCAACCAGCACGTCAACCA 484

Qy       2121 GGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCA 2180

Db        485 GCACCTCGACGTCGACGACACCAAGCACAACCACGTCCACAACCACTTCGACGACACCGA 544

Qy       2181 GGCCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAA 2240

Db        545 CGACTACTAGCTCGAGCACGTCAACCAGCACGTCAACCAGCACGTCGACCAGCACATCGA 604

Qy       2241 GAAGGAGAAGGTGTACCTGAGCTGGGTGCCCGCCCACAAGGGCATCGGCGGCAACGAGCA 2300

Db        605 CGTCGACGACACCAAGCACAACCACGTCCACAACCACTGCGACGACAGCGACGACTACTA 664

Qy       2301 GATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCCTGGACGGCATCGATG 2359

Db        665 GCTCG---AGCACGTCAACCAGCACGTCAACCAGCACCTCGACGTCGACGACACCGAGG 720


RESULT 13
BM320864
LOCUS       BM320864                1132 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM320864
VERSION     BM320864.1  GI:18055270
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1132)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1132   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1132
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
```

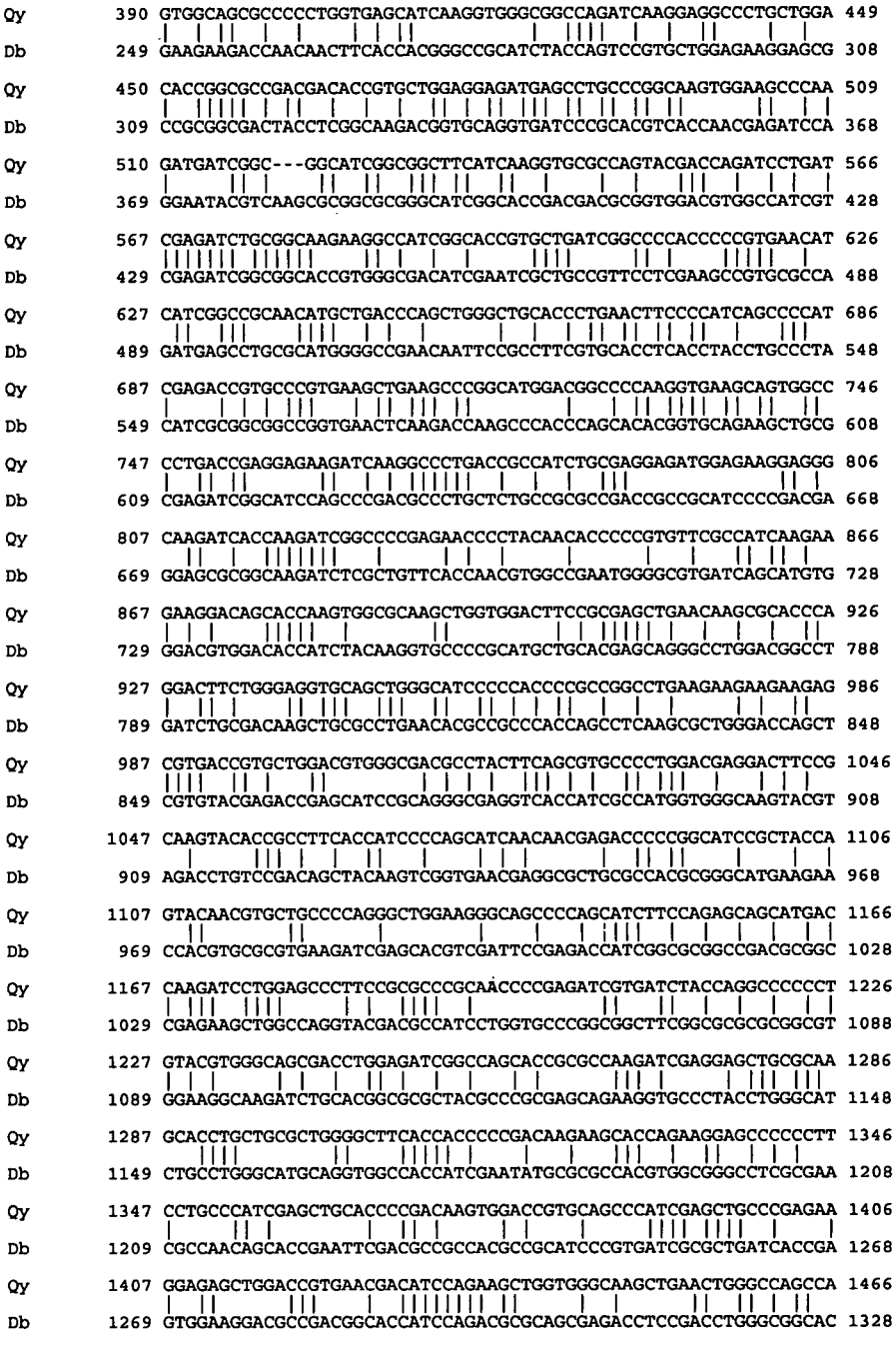
```
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.3%;  Score 80.2;  DB 3;  Length 1132;
  Best Local Similarity   45.3%;  Pred. No. 0.00039;
  Matches  375;  Conservative    0;  Mismatches  444;  Indels    9;  Gaps    2;

Qy        139 CGCGCCCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGAC 198
              ||| |    |||||| || || | |    | ||      || |  ||| | | |||  |
Db         79 CGCCGCGAGGGCAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAG 138

Qy        199 TGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCC 258
              | || |   | | ||  |  |||| ||      | ||   |  |  |  || ||  | |
Db        139 TACAACAGCCCCAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTC 192

Qy        259 CGCGAGTTCCCCAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTG 318
              || || ||  ||  ||   ||| |  || | ||  |  |  |  |  ||       |
Db        193 TGCCAGATCGCCTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCAC 252

Qy        319 CGCGGCGACAACCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCC 378
              |  |  |  | ||  | ||  ||||        | | | | ||    ||    |
Db        253 GAGCTCACCCGCTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACT 312

Qy        379 CAGATCACCCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAG 438
              |  |||  |  ||   |||  ||  |||| | | |  | ||  | |||
Db        313 GGCCTGCTGCTGGCCCGCCGTGTGCTGAAGAAGCTCAACCTCGACTCCAAGTACGAGGGT 372

Qy        439 GCCCTGCTGGACACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAG 498
              | |  |  || || ||||| ||| ||| |||   ||||          | |||| ||
Db        373 GTCAAGAAGGTCAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGT 432

Qy        499 TGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAG 558
              |  ||| ||  | |   || |||| | | | ||  |  | |    ||      |
Db        433 TTCAAGGCCCTGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCC 492

Qy        559 ATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCC 618
              |||  |  |  || |||| |      | || || | ||   |  | ||   ||  | |
Db        493 GCCCTCAAGGGCATGTGCGACGGCGGCGTCAACGTCCCC---CACAGCGAGACCCGCTTC 549

Qy        619 GTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATC 678
              ||   | ||| ||||  ||| | |  |  | |    || || | |  |  |  |||
Db        550 GTCGGCTTCAACGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTC 609

Qy        679 AGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAG 738
              ||  |  ||   ||| |  | |||||||       |    ||||| | ||    |  |
Db        610 GGCGGCCACGTCGCCGCGTACATGAAGCTCCTCAAGGAGCAGGACGCCGCCGCCTTCGAC 669

Qy        739 CAGTGGCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAG 798
              |    |  |  | | | |||      | |    ||| |||   |||| ||||  |
Db        670 CGCCAGTTCTCGCGCTACGCCAAGGAGGGTGTCACCGCCGACATGCTCGAGAAGATCTAC 729

Qy        799 AAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCGCC 858
              |  |||| | |  |     |  || | |  | |  | |  | || | |||  |
Db        730 ACCGAGGCCCACAAGCAGATCCGCGCCGACCCGACCTTCGTCCCCAAGCGCGCCTCGAAG 789

Qy        859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAG 918
              |  |||      |   ||   | ||| ||||||      |       | |   | ||||
Db        790 CCCGAGGGCGCCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAG 849

Qy        919 CGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCC 966
                |  |   | |     | ||||  |  ||||||  |||| || || ||
Db        850 AACCGCGTCGCCCAGAAGAAGGTCCGCCTGGGCTACCCCGACGCCCCC 897


RESULT 14
BM320900
LOCUS       BM320900                1165 bp    mRNA    linear   EST 03-JAN-2002
DEFINITION  rockefeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library
            Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
            sequence.
ACCESSION   BM320900
VERSION     BM320900.1  GI:18055306
KEYWORDS    EST.
SOURCE      Mastigamoeba balamuthi
  ORGANISM  Mastigamoeba balamuthi
            Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE   1  (bases 1 to 1165)
  AUTHORS   Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
            Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
            Philippe,H.
  TITLE     The analysis of 100 genes supports the grouping of three highly
            divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
   PUBMED   11830664
COMMENT     Contact: Muller Miklos
            Laboratory of Biochemical Parasitology
            The Rockefeller University
            1230 York Avenue, New York, NY 10021, USA
            Email: mmuller@rockvax.rockefeller.edu
            Insert Length: 1165   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1165
                     /organism="Mastigamoeba balamuthi"
                     /mol_type="mRNA"
                     /strain="ATCC 30984"
                     /db_xref="taxon:108607"
                     /clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
                     /note="syn: Phreatamoeba balamuthi"
ORIGIN

  Query Match             3.2%;  Score 79.6;  DB 3;  Length 1165;
  Best Local Similarity   45.1%;  Pred. No. 0.0005;
  Matches  428;  Conservative    0;  Mismatches  506;  Indels   16;  Gaps    3;

Qy         90 CATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCCG 149
              | |||||  |      | |  ||||     ||| | |    |   |||||  |     |
Db         46 CGTCAAGAACAAGGCGTACTTCAAGCGCTTCCAGACCCAGTTCCGTCGCCGCCGCGAGGG 105

Qy        150 CAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCG 209
              ||||| || || | |    | ||      || |  |||  | | |||  | | || |   |
Db        106 CAAGACGGACTACCGCGCGCGCCANCTGGTGATCCAGGACAAGAACAAGTACAACAGCCC 165

Qy        210 CCAGGCCAACTTCTTCCGCGAGGACCTGGCCTTCCCCCAGGGCAAGGCCCGCGAGTTCCC 269
              | ||  |  |||| ||       | ||     |  |  | || ||  | | || || || |
Db        166 CAAGTACCGCTTCGTC------GTCCGCTTCACCAACAGGGACATCGTCTGCCAGATCGC 219

Qy        270 CAGCGAGCAGAACCGCGCCAACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGGCGACAA 329
              |  ||   ||| |  || | ||  |  |  |  |  ||          |  | |
Db        220 CTACGCCAAGATCGACGGCGACCACATCCTCGCCGCCGCCTACTCGCACGAGCTCACCCG 279

Qy        330 CCCCCGCAGCGAGGCCGGCGCCGAGCGCCAGGGCACCCTGAACTTCCCCCAGATCACCCT 389
              |  | ||  | ||  ||||         | | | | ||    ||      |    ||
Db        280 CTTCGGCGTCAAGCTCGGCCTGACCAACTACGCCGCCGCCTACGCGACTGGCCTGCTGCT 339

Qy        390 GTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGA 449
              |  |  ||   ||| ||  |||| |  |  | ||  | |||  | |  | ||
Db        340 GGCCCGCCGTGTGCTGAAGAAGCTCAACCTCTACTCCAAGTACGAGGGTGTCAAGAAGGT 399

Qy        450 CACCGGCGCCGACGACACCGTGCTGGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCCAA 509
              || |||||  ||| ||| |||   ||||         |  |||| ||   |  | ||||
Db        400 CAACGGCGAGGACTACAACGTCGAGGAGCTCGACGACGGGCCCCGCCGTTCAAGGCCC-- 457

Qy        510 GATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCCAGTACGACCAGATCCTGATCGA 569
              || ||| ||  |  |   | | ||  |  |  |  ||    ||      |   || |
Db        458 --TGCTCGACGTCGGCCTGGTCCGCACCTCGACTGGCGCCCGCGTGTTCGCCGCCCTCAA 515

Qy        570 GATCTGCGGCAAGAAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCCGTGAACATCAT 629
              |  |   || |  | ||| |  |  |  | ||  ||  |  |||   | ||||
Db        516 GGGCATGTGCGACGGCGGCGTCAACGTCCCCCACAGCGAGACCCGCTTCGTCGGCTTCAA 575

Qy        630 CGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGA 689
              ||||  ||| |  |  |  |  |    |  | ||  || | |  | ||| ||  | ||
Db        576 CGGCGACAAGAAGGAGCTCAACGCCGCGGTCCTCCGCAAGTACATCTTCGGCGGCCACGT 635

Qy        690 GACCGTGCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGGTGAAGCAGTGGCCCCT 749
```



RESULT 15
CL982770
LOCUS CL982770 1680 bp DNA linear GSS 21-SEP-2004
DEFINITION OsIFSC049024 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL982770
VERSION CL982770.1 GI:52420015
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 1680)
AUTHORS Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES Location/Qualifiers
source 1. .1680
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Query Match 3.2%; Score 79.6; DB 10; Length 1680;
Best Local Similarity 41.9%; Pred. No. 0.00051;
Matches 612; Conservative 0; Mismatches 844; Indels 6; Gaps 2;





```
Qy       1467 GATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCGGCGCCAAGGCCCT 1526
             ||   ||  |||     |       | | |   |         ||  |    | ||  |
Db       1329 CATGCGCCTGGGCGCGCAAAGCTCCGACGTCGCCTCCGGCACGCTGGCGCACGAGATCTA 1388

Qy       1527 GACCGACATCGTGCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGAT 1586
               | |||||||  ||  |  | |       |  |   |||  |  || ||||  |
Db       1389 CGGCAACATCGTCACCGAGCGCCACCGCCACCGGTACGAGGCGAACGTGAACTACCTCGA 1448

Qy       1587 CCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCAGCAAG---GACCTGGTGGCCGA 1643
             || || |       | |  |  ||||  || ||  | ||   ||   || | | |  ||||
Db       1449 CCAGCTCCGCAAGGCGGGCCTCGTGATCTCCGCGCTCACGCAGCGCGAACAGCTCACCGA 1508

Qy       1644 GATCCAGAAGCAGGGCCACGACCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAA 1703
             ||||    ||| |   | |||   |    | |     |||  |  | ||  | |    |
Db       1509 GATCGTCGAGCTGCCGCGCGAGGTGCATCCGTGGTTCATCGGCGTGCAGTTCCACCCCGA 1568

Qy       1704 CCTGAAGACCGGCAAGTACGCC 1725
               | ||| ||  |  ||  | |
Db       1569 GTTCAAGTCCACCCCGTGGGAC 1590
```

Search completed: December 31, 2005, 02:31:32
Job time : 8682.48 secs